The Prehistory of Language from the Perspective of the Y-Chromosome

A research guide for linguists from the Genetic-Linguistic Interface Project

Dr. Michael St. Clair, PhD
Trueffelweg 2
70599 Stuttgart, Germany

Email: mstclair@genlinginterface.com
Personal Website: https://genlinginterface.com/

Preface

Many have gone down the wrong path. In the last five years the quest to explain the correlation between genetic and linguistic diversity has employed a methodology called palaeogenomic modeling. Such models were published in prestigious science journals including *Nature*. They have also been reported in mainstream media such as the BBC. Furthermore, the metrics data reflect that they are cited frequently in scholarly journals. These models, however, are flagrantly inconsistent with the archaeological record. They employ the wrong genetic marker and not enough data. I strongly believe that the palaeogenomic modeling “fad” will soon dissipate because of these deficiencies. My research stands ready to yield desperately needed models of language prehistory that are highly reliable. Researchers will have, for the first time, a robust methodology for exploring the correlation between genetic and linguistic diversity.

Table of Contents for Research Guide

PAGE

Chapter 1: Language Prehistory and the Y-Chromosome.

Chapter 2: Haplogroup A.

Chapter 3: Haplogroup B-M60.

Chapter 4: Haplogroup D-M174.

Chapter 5: Haplogroup E-M96.

Chapter 6: Haplogroup C-M130.

Chapter 7: Haplogroup G-M201.

Chapter 8: Haplogroup H-M2713.

Chapter 9: Haplogroup I-M170.

Chapter 10: Haplogroup J-M304.

Chapter 11: Haplogroups L-M20 and T-M184.

Chapter 12: The KR-M526 Paragroup.

Chapter 13: Haplogroups M-P256 and S-B254.

Chapter 14: Haplogroup N-M231.

Chapter 15: Haplogroup O-M175.

Chapter 16: Haplogroup Q-M242.

Chapter 17: Haplogroup R-M207.

Chapter 18: Recommendations, Observations and Future Research.

Chapter 1: Language Prehistory and the Y-Chromosome.

Section 1. Introduction and Organizational Overview.

The correlation between linguistic and genetic diversity is rather straightforward: we inherit our genes and the mother tongue from our parents (for a more detailed discussion, see Cavalli-Sforza 2000). Far more problematic is finding a genetic marker that explains this relationship without complicated statistical analysis. The human Y chromosome overcomes this handicap because it is not subject to recombination. This, in turn, facilitates a high-resolution presentation of the human genetic data that builds highly empirical and highly transparent models of language prehistory. High resolution stems from the ability of the Y-chromosome to delivers hundreds of informative mutations that differentiate the genetic history of one human population from another. Empirical and transparent means that Y-chromosome data are amenable to triangulation with other data sources, such as the archaeological record, the climatological record, and language variation. *Triangulated Y-chromosome based modeling*, in turn, produces far more reliable models of language prehistory because it is built from a convergence of several independent lines of evidence.

This monograph examines the prehistory of language from a Y-chromosome perspective. The goal of this examination is to produce a body of knowledge that will eventually yield a methodology for employing *triangulated Y-chromosome based modeling*. Chapter 1 serves as the introduction. Chapter 18 is the conclusion. Chapters 2 to 17 are the so-called "haplogroup reports." Supplementary figures and data tables have also been prepared. These documents can be accessed from the Genetic-Linguistic Interface webpage. https://genlinginterface.com/

Section 2. Historical Overview of the Search for Informative Molecular Markers.

Cross-disciplinary collaboration between geneticists and linguists has, in fact, long-standing historical precedent. The development of genetic theory began in 1859 with the publication of *On the Origins of Species* by Charles Darwin, a book that sought to explain variation found in the natural world. In 1863, August Schleicher, a giant in the field of historical linguistics, published an open letter to a professor in Jena, Germany. In the letter Schleicher stressed that linguistics and Darwinian Theory represent complementary methodologies. One idea that surfaced repeatedly was taxonomic relationships, meaning that over time languages and organisms evolve from a common ancestor. This view of language diversity prevails today in historical linguistics which utilizes tools such as the comparative method to demonstrate how several languages diverged from a common ancestral language (e.g., Trask 1996). Another interesting idea that surfaced in Schleicher's paper is that factors affecting genetic variation also affect linguistic variation. A simple contemporary example would be the Khoi people of southern Africa and the Waorani people of the Amazon rainforest, who possess significant linguistic and genetic differences because of geographic isolation from each other. Portuguese and Spaniards, on the other hand, exhibit far less genetic and linguistic variation because of close geographic proximity.

The term "marker" refers to a section of DNA. Geneticists have found several different polymorphic markers for measuring genetic variation among human populations. The term "polymorphic" means that a section of DNA can vary from one person to the next. Polymorphic protein markers used to assess human genetic variation are generally referred to as "classical markers" in the literature. In 1919, Ludwik and Hanka Hirschenfeld, two researchers at a military hospital, published a study that used classical markers for the first time to assess human variation. The researchers utilized ABO blood groupings to find patterns of variation among different nationalities and ethnic groups. In 1994, Cavalli-Sforza, Menozzi, and Piazza published the most comprehensive study of human variation based on classical markers. However, the study conceded (9-10) that another type of marker, mitochondrial DNA (mtDNA), would be a better choice for population studies, but at the time a sufficient number of haplogroups had not yet been discovered.

During the 1980's improved sequencing technology enabled geneticists to focus on the nucleotide bases that form the rungs of the DNA molecular ladder. From this development Brown (1980) identified mitochondrial DNA as a polymorphic marker. About five years later, in 1985, Casanova and others identified the non-recombining region of the Y-chromosome as polymorphic markers in human populations. That same year Hill and others (1985) published the first autosomal marker study. From about 1985 onwards, hundreds of reports have been published by geneticists to detail human autosomal, mtDNA and Y-chromosome DNA variation throughout the world. From a macro-perspective, molecular genetic evidence places the origins of *Homo sapiens* in Africa. Cann, Stoneking and Wilson, for example, published a paper in 1987 asserting that female human beings trace their genetic history to a woman living in Africa about 200,000 years ago, the so-called "mitochondrial Eve." African origins for the so-called "Y-chromosome Adam" emerged several years later using data from the non-recombining region of the Y-chromosome (Underhill et al. 2000).

Section 3. Important Key Concepts.

3.1. Overview.

Anthropologists, archaeologists and geneticists have recognized the potential of Y-chromosome data as a tool for deciphering the human past. However, the potential of this data source as research tool for linguists remains largely unexplored. Accordingly, this present section offers several important concepts related to the human Y-chromosome with the goal of persuading linguists that the Y-chromosome data are a useful tool for investigating the prehistory of language. Additionally, this discussion helps to define why haplogroups are utilized as the organizational backbone for presenting a multi-disciplinary synthesis of several data sources that can build *triangulated Y-chromosome based* models of language prehistory. Organization according to geographic regions or language classification represents two viable alternatives. However, organizing the discussion into the unique divisions of human genetic data that are represented by haplogroups provides, arguably, the most elegant summation of the data.

3.2. Recombination.

In order to understand how the Y-chromosome behaves differently from other genetic markers, it is necessary to briefly discuss Mendelian genetics, which is often part of high school and introductory college biology instruction. According to Mendelian genetics we inherit our genes from both parents. However, the Y-chromosome plays by its own genetic rules in that it is only passed from a man to his son. The Y-chromosome is one of the two sex-chromosomes in the human genetic inventory, or human genome. The other sex chromosome is the X-chromosome. During reproduction, two X-chromosomes yield female offspring, and an X-chromosome and a Y-chromosome yield male offspring. Another "rule" of Mendelian genetics is recombination. During human reproduction, the genetic cards are essentially “reshuffled,” or more precisely, recombination occurs. However, the genetic material contained in the Y-chromosome, for the most part, escapes recombination.

In order to explain how the Y-chromosome avoids recombination, it is necessary to briefly discuss the evolutionary history of this chromosome. The sex-determining locus of the Y-chromosome not only codes for maleness in humans, but in all mammals. This section of the Y-chromosome, however, only represents a fraction of its entire length. During the evolutionary history of mammals, about 300 million years, the Y-chromosome has slowly “degenerated” or degraded (e.g. Lahn et al. 2001). When mammals first evolved, the Y-chromosome “behaved normally” in that the entire chromosome recombined with the X chromosome. Now, as the result of slowly evolving structural decay, about 95% of the entire length of the Y-chromosome has been damaged, emerging in what the geneticists call a “non-recombining region.” This large non-recombining region means that during reproduction very little genetic exchange occurs between the X and Y chromosome. Consequently, the Y-chromosome has been transmitted largely intact from one human male to the next for the last 300 thousand years.

3.3. Mutation.

The Y-chromosome is unique due to uniparental inheritance and the absence of recombination. Consequently, men inherit a large section of the human genome that remains unaltered when the genetic cards are reshuffled. However, the non-recombining region of the Y-chromosome can and often varies from one Y-chromosome to the next. Geneticists describe this variation as mutation. In population studies examining Y-chromosome variation, one type of a particularly informative mutation is defined by single nucleotide polymorphisms.

In order to better understand the concept of single nucleotide polymorphisms (or SNP’s) it is necessary to focus on the molecular structure of deoxyribonucleic acid, or DNA. The molecular “ladder” of DNA has “rails” formed by alternating sugar and phosphate molecules. The “rungs” of this ladder, known as nucleotides, are formed by bonding two molecules having a nitrogenous base; either adenine and thymine, or guanine and cytosine. Since the non-recombining region of the Y-chromosome has about 60 million molecular “rungs,” or base pairs, geneticists have a vast region of genetic information to harvest the evolutionary history of *Homo sapiens*.

The order of bonding along the Y-chromosome molecular ladder alternates, meaning the nucleotides appear in one of four different combinations: adenine/thymine, thymine/adenine, guanine/cytosine and cytosine/guanine. A single nucleotide polymorphism occurs when one of the rungs of this molecular ladder changes, or mutates. Mutations occur when a nucleotide is substituted for another, or when nucleotides are added or deleted. So, for example, the recently discovered R1b-DF27 mutation involves the substitution of a guanine/cytosine bond with an adenine/thymine bond at position 21380200. It should be noted that mutations are extremely rare. In fact some early Y-chromosome studies (e.g. Underhill et al. 2001) initially referred to these Y-chromosome mutations as “unique event polymorphisms” because they are so rare they only occur once during human evolution.

At this point it is important to emphasize a concept known as neutral selection. Those who have taken an introductory biology or physical anthropology course have probably encountered the term “natural selection,” initially proposed by the Charles Darwin. This theory accounts for different animal and plant species based on fitness, or survival of the fittest. According to this theory differentiation among species arose as the result of a mutation that enabled the plant or animal to survive in a given environment long enough to pass on its genes to the next generation. Y-chromosome mutations, however, are classified as selectively neutral, meaning they do not confer any reproductive advantage. Likewise, these mutations are not disadvantageous. Introductory biology courses often emphasize that genetic mutations can be harmful or fatal to living organisms. For example, among humans one of the most recognized harmful genetic mutations is sickle cell anemia. In contrast to sickle cell anemia and other harmful genetic mutations, Y-chromosome mutations are benign. This explains, partially, why Y-chromosome mutations survive while many genetic mutations affect reproductive success and are consequently eliminated from the gene pool.

3.4 Nomenclature and Phylogenetic Relationships.

As explained in the above paragraph, single nucleotide polymorphisms are a mutation found in the non-combining region of the Y-chromosome. Geneticists comb the non-combining region of the Y-chromosome to identify these mutations. The presence or absence of mutations, or single nucleotide polymorphisms, can distinguish the genetic history of one population from the next. Y-chromosome single nucleotide polymorphisms are generally described as paragroups, haplogroups, or sub-haplogroups.

Haplogroups are identified using a nomenclature system first standardized in 2002 by the Y Chromosome Consortium (YCC). This nomenclature system identifies haplogroups by cladistic name and mutation. The cladistic name utilizes set theory and assigns an uppercase letter to identify one of the current twenty different major haplogroups. Then, to identify variants of the major haplogroups, or sub-haplogroups, the uppercase letter is followed by a combination of numbers and lower-case letters.

Besides the cladistic name, YCC 2002 recommended the addition a mutation number to the cladistic name of single nucleotide polymorphism. This number is preceded by a letter such as "M," or "P" or "V." These letters generally identify the laboratory that discovered the mutation. For example, one very common haplogroup found in Europe is I-M170, the "I" meaning haplogroup I, and M170 referring to mutation number 170, which was discovered by Peter Underhill at Stanford University (hence "M"). An example of a sub-haplogroup is the I1-M253 mutation, commonly found in Scandinavia. The "1" is used to classify I1-M253 as a sub-haplogroup of haplogroup I-M170. Often the literature does not make a formal distinction between haplogroups and sub-haplogroups, and thus I-M170 and I1-M253 would simply be reported as "haplogroups." Additionally, since the methodology used to build these hierarchical relationships is called cladistics, the terms "clade" and "subclade" are sometime used to label haplogroups and sub-haplogroups.

As noted above, the initial effort to standardize the Y-chromosome nomenclature was in 2002 and the paper published by the Y-Chromosome Commission. Since 2002 important updates to the nomenclature have included Karafet et al. 2008, Oven et al. 2014, Karmin et al. 2015, Karafet 2015, and Poznik et al. 2016. Additionally, since 2002 geneticists have identified thousands of subhaplogroups (e.g. R1b-DF27) that contribute to our current picture of global Y-chromosome variation. An updated list of these sub-haplogroups is maintained on the website (https://isogg.org/) of the International Society for Genetic Genealogy (ISOGG).

Since 2002 the Y-chromosome tree has grown quite large and the cladistic names of mutations have become quite long and cumbersome. For example, the R1b1a1a2a1a2a-DF27 mutation describes a sub-haplogroup that is found on the Iberian Peninsula. Consequently, in practice the nomenclature would be shortened to R-DF27, or perhaps R1b-DF27. Additionally, geneticists sometimes publish papers that omit the mutation numbers, and present data simply with the cladistic identifier (e.g. R1b instead of R1b-M343). This practice is problematic because cladistic identifiers change frequently whereas mutation numbers remained much more stable. An example would be the O3-M122 mutation (e.g. Karafet et al. 2008) which is now O2-M122 (ISOGG 2017).

The ISOGG website has become the repository of phylogenetic updates for Y-chromosome mutations. However, the focus of this organization is genealogical research, and as such, many of the polymorphisms listed by the organization are not informative markers for linguistic research. Thus the task for linguists is to identify linguistically informative mutations. Additionally, the linguist must keep abreast of the phylogenetic updates posted on the ISOGG website. Finally, linguists have to convert old nomenclature to the new standard (e.g. O3-M122 to O2-M122) so that old data can be compared to new data.

At this point the reader is directed to Supplementary Figure 1.1 which presents the mutational steps from Y-Chromosome Adam to the main haplogroups (please note that the diagram expands onto a second page). The figure depicts hierarchical relationships among Y-chromosome mutation. Y-chromosome Adam at the top of the diagram represents the evolution of *Homo sapiens*. At the bottom of the diagram are the main haplogroups: A (Adam), B-M181, D-M174, E-M96, C-M130, G-M201, H-M2713, I-M170, J-M304, L-M20, T-M184, M-P256, S-B254, N-M231, O-M175, Q-M242, R-M207, and KR-M526*. These haplogroups represent unique segments of contemporary Y-chromosome diversity. Between Y-chromosome Adam and the haplogroups are several important "paragroups" that help to decipher the evolutionary history of Y-chromosome diversity. For example, the DR-M168 paragroup resents the genetic ancestor of all the haplogroups that evolved outside of Africa.

At this time a discussion of several points are in order. First, Figure 1.1 deviates *slightly* from the YCC 2002 and ISOGG 2017 standard. We label the F-M89 and P-P295 mutations as higher level paragroups rather than haplogroups. Secondly, we avoid the use of "haplogroup" to describe a "paragroup" or "sub-haplogroup." These terms represent very important distinctions that some genetic studies fail to make. Thirdly, linguistically informative mutations are generally found downstream from the main haplogroups. The internal structure of the haplogroups will be described in due course in Chapters 2 to 17. Fourthly, we insist that a mutation carries the mutational number (e.g. M343 for the R1b). Many studies only carry on the cladistic identifier (such R1b for the M343 mutation) which makes the task of converting from old nomenclature to new nomenclature extremely difficult. Finally, we navigate the discussion of hierarchical relationships using the terms "downstream" and "upstream." For example, the R1b-M343 mutation is downstream from the R-M207 haplogroup. Upstream from the R-M207 haplogroup is the P-P295 paragroup.

Linguists should note the phylogenetic mapping of genetic relationships is akin to mapping linguistic relationships with tree-like language family diagrams. English, for example, is part of the West Germanic sub-branch of the Germanic branch of the Indo-European language family.

3.5. The Dating of Mutations.

The decipherment of hierarchical relationships within non-recombinant region of the Y-chromosome is aided by the ability to determine, at least roughly, when Y-chromosome mutation evolved. For example, haplogroups D-M174, E-M96 and C-M130 diverged from the DR-M168 paragroup about 70 thousand years ago (Poznik et al. 2016; Supplementary Table 10). Dating methodologies used by geneticists to make these estimates are indeed very complex and involve attempts to determine an average rate of mutation for Y chromosome polymorphisms (e.g. Balanovsky 2017a). Beginning with the study by Zhivotovsky et al. (2004), several studies have sought to refine this methodology. Initially, the methodology utilized a type of mutation called short tandem repeats. Today Karmin et al. (2015) and Poznik et al. (2016) present the latest developments in this ongoing effort. They are able to present more robust estimates by utilizing whole genome sequencing. Moreover, Poznik et al (2016) is particularly significant for two reasons. First, the study not only utilizes whole genome sequencing, but also attains it conclusions from ancient DNA found in Romania and Siberia. The dating results, as presented in Supplemental Table 10 of their study, correlate well with the archaeological record.

3.6. Phylogeography.

The reader is invited to review the second page of Supplementary Figure 1.1 which provides a short summary of where the various Y-chromosome haplogroups are found. For example haplogroup D-M174 is found in East Asia and haplogroup S-B254 is found in Australia. Furthermore, unlike autosomal markers, it is fairly easy to generate phylogeographic maps of Y chromosome mutations. Such maps trace the frequency of a given mutation across geographical distance. The ability to analyse data in this matter, and the ability to estimate when a mutation evolved, helps researchers to determine how and when a prehistoric group may have migrated.

Needless to say a migration consists of a geographic point of origin and a geographic point of termination. Sometimes the point of origin has the greatest frequency of a particular haplogroup and over distance the frequency of this haplogroup diminishes. For example, the J2a-M67 mutation arose in the Near East during the Mesolithic. During the Neolithic this group migrated to Western Europe, and along this route the frequency of the mutation decreased because of admixture with other groups already living in the new territory (see Chapter 10). However, some prehistoric migrations show an opposite pattern or cline of haplogroup frequencies, where the point of origin has the lowest frequency of a certain haplogroup, and the terminal end of the migration has the highest. The I1-M253 mutation, for example, potentially represents a prehistoric migration from the current Spanish/French border to Scandinavia. Along this route the frequency of I1-M253 increases, perhaps because a prehistoric group moved into unoccupied territory, or perhaps the group acquired a novel survival strategy that gave them a reproductive advantage (see Chapter 9: Section 4).

3.7. Population History.

Genetic differences between populations help to decipher demographic milestones of the human prehistory. For example, the distribution and frequency of the Q1b-M3 mutation help to decipher the human colonization of the Americas (see Chapter 16). From a very clinical point of view the term “population” refers to a group of potentially inter-breeding individuals. In practice several factors influence how people choose a partner with whom they eventually have children. These factors may include ethnicity, religion, or socio-economic status. However the most salient factor for the purposes of this discussion is geographic distance. For example southern Africa and the Amazon rainforest are obviously separated by vast geographical distance of several thousand kilometers. As a result the

Khoisan people of southern Africa and the Waorani people of the Amazon rainforest possess genetic differences because of geographic isolation from each other. From a Y-chromosome perspective this explains why the Khoisan of southern Africa have the A-M6 mutation and why the Yanomami of South America have the Q-M3 mutation; and why the Khoisan do not have Q-M3 mutation and why the Yanomami do not have haplogroup A-M6 mutation.

The above discussion of how populations have become isolated from each other conforms to a demographic model called "genetic drift." This concept suggests that a leveling of genetic diversity occurs among small isolated populations, a demographic scenario that characterizes most of the human prehistory. In order to explain how geographic distance has separated populations throughout the world it is necessary to turn to human prehistory and with that, the archaeological record of human migrations. Archaeological and genetic data place human origins in Africa about 300 thousand years ago (see Chapter 2). Around 100 thousand years humans left Africa. Between 40 and 50 thousand humans colonized Europe, South Asia, East Asia, Papua New Guinea, and Australia (see Chapter 4). About 15 thousand years ago people crossed over the Bering land bridge from Asia into North America (see Chapter 16).

The term "founder effect" illustrates a type of genetic drift and is a useful concept in understanding the development of genetic differentiation between populations. Founder effect describes a situation whereby mutations frequencies are altered when a group of people separates from a larger population. For example, Austronesian-speakers on Papua Guinea have a mixed ancestry of East Asian and Melanesian Y-chromosome mutations. A sub-population then carried Austronesian languages across Oceania about two thousand years ago. As the result of numerous founder effects, East Asian mutations disappeared once the expansion had reached Rapa Nui (or Easter Island). See Chapter 15: Section 7 for additional information.

Another type of genetic drift that sometimes contributes to genetic differentiation between populations is "bottleneck." This term describes a situation where perhaps disease or a natural disaster suddenly reduces the size of a population that is isolated and relatively small. This sudden reduction in population reduces the amount of haplogroup variation, and like founder effect, accelerates drift. For example, the Y-chromosome data suggest that the Toba volcano explosion may have produced a bottleneck effect among human populations roughly 70 thousand years. Additionally, the data suggest genetic variation in Paleolithic Europe was characterized by that I-M170 and C-M130 haplogroups, as well as the NO-M214 paragroup. Today, the only remaining founder mutations among contemporary European is variants of the I-M170 haplogroup. It appears as though the Last Glacial Maximum produced a bottleneck effect in Europe roughly 20 thousand years ago that reduced the size of human populations in Western Europe. Perhaps the bottleneck was caused by a reduction in the number of reindeer and with that, a shortage of food (see Chapter 17: Section 6).

3.8. Advantageous of the Y-Chromosome over Other Molecular Markers.

Again, the term "marker" refers to a section of DNA. Common molecular markers used for human population history are autosomal DNA, mitochondrial DNA (mtDNA) and the non-recombining region of the Y chromosome. As such, one potential criticism of this resource guide is that it focuses almost exclusively on Y-chromosome variation and excludes autosomal and mtDNA perspectives. The counter-argument is that the Y-chromosome provides a perspective of human genetic history that has high-resolution and transparency.

Mitochondrial DNA data, on the other hand, lack resolution and autosomal markers lack transparency. Indeed, mitochondrial DNA has many of the desirable features of the non-recombinant region of the Y-chromosome, such as the absence of recombination and the ability to order mutations within a phylogenetic tree. However mtDNA data are gathered from a small section of the human genome that has only 16 thousand base pairs. The non-recombinant region of the Y-chromosome, on the other hand, has 60 million base pairs. As such the Y-chromosome offers a much more resolved picture of human prehistory. For example, mtDNA haplogroups lack counterparts for Y-chromosome R1a-M420 and R1b-M343 mutations as well as the N-M241 haplogroup. The comparison between Y-chromosome and mtDNA data is analogous to the picture quality one obtains from a two megapixel camera versus a ten megapixel camera.

Turning now to autosomal studies, one potential benefit of this marker is the ability to obtain a genetic perspective for both genders, whereas the perspective of mitochondrial DNA is arguably for females and the Y-chromosome perspective is arguably for male. However data from autosomal markers are affected by recombination and thus require incredibly complex statistical analysis. On the other hand, the two uniparental markers, mitochondrial DNA and the Y-chromosome, can be analyzed without complicated statistical methodologies. Moreover, since autosomal data focuses primarily on the frequency of alleles rather than the presence or absence of mutations, autosomal data are not amenable to analysis by means of phylogenetic trees, whereas this is possible with mtDNA and the Y-chromosome (cf. Oven and Kayser 2008; Oven et al. 2014).

One perceived problem with Y-chromosome data is that they deliver a picture human prehistory that is biased toward the male gender. In practice, the application of this tool for linguistic research has not confirmed such bias. Rather, this section of the human genome simply acts as an effective "trap" that captures important human demographic milestones that decipher language prehistory for both genders. For example, mtDNA and Y-chromosome perspectives place human origins in Africa (e.g. Oppenheimer 2012).

The huge disadvantage associated with Y-chromosome data (e.g. Jobling and Tyler-Smith 2003) is "ascertainment bias." When investigators collect samples from a population, a small number of samples may skew the actual frequency of a mutation within the population. A larger number of samples, on the other hand, achieve a more realistic picture of genetic diversity. Extending this argument further, more samples yield a more resolved model of language prehistory. Thus, for example, a well resolved picture of language prehistory is available for Indo-European, whereas the picture is highly ambiguous for Eskimo-Aleut and Eyak-Athabaskan (cf. Chapters 10 and 16).

Section 4.0. Conclusions.

Cross-disciplinary cooperation between geneticists and linguists has, in fact, long-standing historical precedent. Chapters 2 to 17 build upon August Schleicher's ideas by presenting a Y-chromosome perspective of language prehistory. The goal of this examination is to produce a body of knowledge that will eventually yield a methodology for employing *triangulated Y-chromosome based modeling*.

Chapter 2: Haplogroup A.

Section 1. Overview.

All of the current human Y-chromosome variation in the world evolved from the theoretical Y-Chromosome Adam. Unlike the other main haplogroups, haplogroup A is defined by an individual rather than a mutation. According to Mendez et al (2013) Y-chromosome Adam evolved in west central Africa about 338 thousand years ago. This conclusion stems from genetic data collected from an African American having the oldest known haplogroup A lineage (A00-AF6/L1284) as well as variants of this mutation as found among the Mbo people of western Cameroon. Hublin et al. (2017) published a study that provides archeological support for this position based on fossil remains found at Jebel Irhoud in Morocco. Specifically, the archaeological data from Hublin and others generally agrees with genetic data from Mendez and others (2013) with respect to the location of human evolution (western Africa) and when humans evolved (roughly 300 thousand years ago). The report by Hublin is especially significant because previous fossil evidence (e.g. White et al. 2003) had placed human origins in eastern Africa about 200 thousand years ago, which in turn, undermined the conclusions reached by Mendez and others in 2013.

At this point the reader is directed to the top of Supplementary Figure 2.1 which provides a phylogenetic overview of the important mutations within haplogroup A. According to Poznik et al. (2016), A0-V148 and A1-V168 split from A00-AF6/L1284 about 190 thousand years ago, and A1a-M31 and BR-M42 separated from A1-V168 about 160 thousand years ago. Moving now towards the bottom of Supplementary Figure 2.1, the reader finds three mutations that encompass almost all of the current haplogroup A variation in Africa and the world: A1b-M6, A1b-M51 and A1b-M13. These mutations have significant regional and linguistic differences that will be detailed in the following two sections.

Section 2. Southern African Khoisan.

As discussed in the previous section, haplogroup A represents the most ancestral and oldest Y-chromosome lineages, having evolved in Africa about 300 thousand years ago. Today haplogroup A is found almost exclusively among populations living on the African continent. As reflected by recently prepared frequency contour maps (Rowold et al. 2016) haplogroup A is concentrated either in southwestern Africa or parts of eastern Africa. Within a southern African context, haplogroup A stands as the genetic relic of pre-agricultural African populations (e.g. Rosa et al. 2007; Batini et al. 2011). Support for this conclusion comes from the so-called Khoisan people, a collection of hunter-gatherer cultures found in this region. Barbieri et al. (2016) found that among the Khoisan the A-M6 mutation attains a frequency of about thirteen percent and the A-M51 mutation attains a frequency of about twenty percent. According to the same study, both mutations seem to be the genetic signature of the Khoisan. The A-M6 mutation was not found in the Bantu populations of the region, and among the Bantus, the A-M51 mutation attains a frequency of less the four percent. Finally, the same study concluded that both mutations are indeed confined almost exclusively to southern Africa.

It should also be emphasized that the term “Khoisan” has both cultural and linguistic components. As noted above, from a cultural perspective the Khoisan are the descendants of pre-agricultural hunter-gatherers who evolved and remained on the African continent. Furthermore, they remained hunter-gathers and resisted assimilation with Bantu farmers who migrated from west central Africa to South Africa beginning about five thousand years ago (see Chapter Five: Section 7 for additional information**)**. Turning now to the linguistic component of the term Khoisan, populations lumped into this category speak languages that fall into one of three language families: Khoe-Kwadi, Kx'a and Tuu. This is quite significant from a linguistic perspective as the Khoisan not only resisted the adoption of the Bantu farming culture, but also resisted shifting to the Niger-Congo languages spoken by the Bantus.

The above explanation of the term "Khoisan" facilitates an important discussion of two critical points that linguists need to know. First, geneticists use the term "Khoisan" ubiquitously in published reports that describe genetic variation in Africa (e.g. Underhill et al. 2000; Tishkoff et al. 2007; Barbieri et al. 2016). Secondly, as noted by Mitchell (2010), the term "Khoisan" seems to lump too many groups into one basket when, in fact, each group potentially has cultural, linguistic and genetic histories that should be evaluated independently. For example, Barbieri et al (2016) aggregated the genetic data for the !Xóõ, ||Ani, ||Gana, and Shua people to report the frequency of haplogroup A among "Khoisan" rather that providing a frequency for each population. This is especially disappointing for linguists because it might be beneficial to compare the genetic history of Tuu languages (e.g. the !Xóõ) with that of Khoe-Kwadi speakers (e.g. ||Ani, ||Gana, and Shua).

Section 3. East African Populations.

As noted in the previous section, populations having a significant frequency of haplogroup A are concentrated either in southwestern Africa or eastern Africa. Populations in southwestern Africa have the A-M6 and A-M51 mutations. However, eastern African populations have the A-M13 mutation. This mutation attains an especially high frequency (fifty percent or greater) among some Nilo-Saharan populations in this region, which include the Dinka, Gumuz and Shilluk. Other Nilo-Saharan speaking populations (e.g.Nubians, Ng'arkarimojong, Maasai, Alur and Masalit) also have a significant frequency of the mutation (see Wood et al. 2005; Tishkoff et al. 2007; Hassan et al. 2008; Gomes et al. 2010; Batini et al. 2011; Haber et al. 2016; and Supplementary Table 2.1 for additional details).

Besides Nilo-Saharan speaking populations, the A-M13 mutation also attains a significant frequency among some Afro-Asiatic speaking populations in east Africa, such as Amhara, Oromo and Welayta (see Wood et al. 2005; Hassan et al. 2008; Batini et al. 2011, Haber et al. 2016 and Supplementary Table 2.2 for additional details). Finally, the A-M13 mutation has a moderate frequency among some Niger-Congo speaking populations (see Wood et al. 2005; Batini et al. 2011 and Supplementary Table 2.3 for additional details). What is particularly interesting about the Niger-Congo data stems from the discovery of A-M13 mutations in population outside of eastern Africa, in central or western Africa. Within central Africa, the only Nilo-Saharan population for which haplogroup A data are available is the Sara people of Chad (Haber et al. 2016). However, among these people, the A-M13 mutation only attains a very small frequency of two percent.

The above paragraph and the discussion of the Sara people underscore a current deficiency in that almost all the genetic data for Nilo-Saharan speakers comes from populations living in eastern Africa. Additional data from Nilo-Saharan populations in central and western Africa, such as the Songhai, may well support a position that places the geographic origins of Nilo-Saharan languages within the Sahel, a transition zone that runs west to east across Africa, separating the Sahara Desert and Sub-Saharan Africa. Such a position certainly agrees with the geographic distribution of Nilo-Saharan languages. As shown by Supplementary Figure 2.2, Nilo-Saharan languages extend eastwards from Mali across the Sahel into Sudan and Ethiopia, and then southwards into East Africa and Uganda, Kenya and Tanzania.

Further support for the Sahel as the geographic point of origin for Nilo-Saharan stems from a 2010 study published by Gomes and others. Here, researchers analyzed A-M13 data and based on this data, estimated that Nilo-Saharan populations arose about 15 thousand years ago. Additionally, the same study determined that within the Nilo-Saharan language family, Eastern Nilotic and Western Nilotic languages diverged and expanded four to six thousand years ago. These dating estimates are significant for defining the Sahel as the geographic point of origin for Nilo-Saharan as they correlate well with climate change that occurred within the Sahel during the early Holocene between 8500 and 5300 BC.

During the Holocene, which began about 12 thousand years ago, warmer temperatures caused the ice glaciers to retreat across the Northern Hemisphere. Further south, the same event triggered a temporary but dramatic change within the Saharan Desert of northern Africa that lasted three thousand years. Normally this area stands as one of the most inhospitable regions of the world, with endless miles of sand dunes, temperatures that approach 50 degrees centigrade, and no rainfall. However about ten thousand years ago, as the result of global climate change, monsoon rains came to the region. During this "humid phase" rain transformed the desert into a savanna ecosystem characterized by grassland and widely spaced trees. Furthermore, the monsoon rain brought rivers and lakes to this region and with that, people.

A fascinating study by Drake et al. (2011) presented the results of satellite imagery data that confirmed the presence of a complex system of rivers and lakes that arose during the early Holocene in the Sahara Desert, something that researchers had long suspected as the result of the archaeological record. Drake and others also discussed the numerous barbed bone points that have been collected by archaeologists within this complex system of rivers and lakes. During the last humid phase in the present-day Sahara Desert, Stone Age people utilized these bone points to make harpoons. This technological adaptation enabled people to harvest the hippos, crocodiles and fish that thrived within the complex system of lakes and rivers. Drake and others provided a map that illustrates where the bone points have been found, which actually encompasses the entire Sahel. The study then suggests that the location of these artifacts offer a good correlation with the distribution of Nilo-Saharan languages. To support this position Drake and others present a list of cognates for "crocodile" and "hippo," as found in several of the Nilo-Saharan languages.

Kuper and Kröpelin in their 2006 study focused on the eastern Sahara and effects of climate transformation as it occurred during the last humid phase in this region. The eastern Sahara encompasses the Western Desert of Egypt, northwest Sudan, parts of Libya and Chad. Interestingly, at the time human populations could not live in the Nile Valley as this region was too "marshy." According to the study, west of the Nile, in the savanna that existed seven to ten thousand years ago, people initially survived by hunting and gathering. People in the region later adopted pastoralism, the herding goats and sheep that came from the Middle East, and cattle that have their origin in Africa. Then around 5300 BC the rain suddenly ended, and region became once again, almost overnight, a desert. As the result of desertification, some populations migrated to Nile Valley, where the region had become, in the meantime, more habitable as the result of dryer climate. Those that settled along the Nile River adopted cereal cultivation and became the founding population of Pharaonic Egypt. Other populations, instead of settling along the Nile, escaped desertification of the Sahara by migrating into the Sudan and later eastern Africa. These populations retained cattle pastoralism.

The above discussion of the 2006 study by Kuper and Kröpelin supports the idea that the current distribution of Nilo-Saharan languages may well be a product of climate change that occurred in the Saharan Desert about seven thousand years ago, when this area reverted from a savannah ecosystem to a dry desert. The origins of this language family seem to be linked to the origins of East African pastoralism. This subsistence strategy stands as the traditional food economy documented among many of the Nilo-Saharan speaking populations, such as the Dinka and Maasai. Nilo-Saharan speakers may have brought cattle pastoralism into eastern Africa when the last humid phase ended in the Sahara, roughly seven thousand years ago.

Another compelling reason to undertake additional Y-chromosome sampling of Nilo-Saharan speaking populations in the Sahel is that haplogroup A data may also clarify whether Niger-Congo and Nilo-Saharan languages evolved from a common ancestral proto-language. This topic was explored by the researcher Roger Blench in an unpublished 2006 paper with the title *The Niger-Saharan Macrophylum*. This possibility seems to be supported by the presence of the A-M13 mutation among Niger-Congo speaking population in the western and central Sahel (Wood et al. 2005: Batini et al. 2011). Among the Tupuri of Cameroon, the frequency of A-M13 is between eleven and twenty-two percent, and among the Igala of Nigeria the frequency is a small five percent.

Section 4. Important Phylogenetic Relationships.

Within Supplementary Figure 2.1 the A-M51 and A-M13 mutations are highlighted to illustrate a point. Both mutation are phylogenetically close, being what the geneticists call "sister clades." As noted earlier, the A-M51 and A-M6 mutations are the genetic signature of Khoisan populations in southern Africa, and the A-M13 mutation is the signature of some Nilo-Saharan populations in eastern Africa. Given the fact that the east African Nilo-Saharans and southern African Khoisan are separated by a distance of roughly four thousand kilometers, one would expect greater genetic distance between the A-M51 and A-M13 mutations. Henn et al. (2011) seem to agree by suggesting that both mutations separated about 40 thousand years ago. Nevertheless, Scozzari et al. (2012) suggest a much more recent split, which seems more consistent with the cladistic relationships. These differing opinions are raised to illustrate a need to perhaps further resolve phylogenetically relationships within haplogroup A. Moreover, the study recently published by Barbieri et al. (2016) reports the discovery of a "new" haplogroup A mutation for which the study does not provide a defining mutation. Based on personal communication with one of the study researchers, this "new" mutation is probably very close to the A-M6 mutation reported by Wood et al. (2005) and Gomes at al. (2010). Thus, the A-M6 mutation, like the A-M51 and A-M13 mutations, also illustrates a need to further refine the phylogeny of haplogroup A.

Section 5. Conclusions.

From a linguistic perspective further refinement of the haplogroup A phylogeny, coupled with additional frequency data for African populations, hold the potential for elucidating the linguistic prehistory of the Khoe-Kwadi, Kx'a, Tuu, Nilo-Saharan and Niger-Congo languages. Khoe-Kwadi, Kx'a and Tuu probably have ancient roots that date potentially to the evolution of human language (see, also, Chapter 3: Section 5). Nilo-Saharan and Niger-Congo, on the other hand, may have split from an ancient Niger-Saharan macrophylum. However, the concept of a Niger-Saharan macrophylum presents a problem that requires linguistic reconstruction, a tool that might not work in this case because of the time depth involved. Alternatively and perhaps far less speculative, the linguistic, genetic and archaeological data simply place the Holocene and the Sahara Desert as the temporal and geographic starting points for a discussion of Niger-Congo and Nilo-Saharan.

As noted earlier (Section 3), a moderate frequency of haplogroup A has been found in some Afro-Asiatic speaking populations in eastern Africa. While Niger-Congo and Nilo-Saharan arose in Africa, the Afro-Asiatic languages of Africa appear to be a Holocene import from Southwest Asia (see Chapter 5: Section 5 and Chapter 10: Section 3).

Chapter 3: Haplogroup B-M60.

Section 1. Overview of Haplogroup B-M60.

Like haplogroup A (see Chapter 2) haplogroup B-M60 arose and remained in Africa. The reader is now directed to Supplementary Figure 3.1 which provides a phylogenic overview of this haplogroup and its important downstream variants. According to Poznik et al. (2016) the B-M60 haplogroup evolved about 100 thousand years ago. B2a-M150 and B2b-M112, the most informative haplogroup B variants, evolved about 50 thousand years ago (Barbieri et al. 2016). Frequency contour maps generated by Rowold and others in their 2016 study reflect that B-M150 attains its highest frequency in central western Africa, and B-M112 is concentrated further south in central Africa. As detailed in this present chapter, the distribution of B-M150 reflects the Bantu expansion of agriculture from central western Africa to southern Africa beginning about three to five thousand years ago and ending about fifteen hundred years ago. B-M112, on the other hand, stands as the genetic signature of African hunter-gatherers. For linguists the B-M150 and B-M112 markers present an opportunity to explore language contact theory and more specifically, agriculture as a vehicle for language shift. Additionally, the B-M112 marker provides an opportunity to explore the history of the so-called "click" languages. Are click consonants the oldest phoneme?

Section 2. Macro-Linguistic and Macro-Cultural Relationships within Africa.

The reader, at this point is directed to the language map of Africa as provided in Supplementary Figure 2.2. The major language families of Africa include Afro-Asiatic, Nilo-Saharan, Niger-Congo, and Khoisan. At this point it is necessary to introduce Hadza and Sandawe, two isolate languages found in Tanzania. These isolates along with Khoisan have click consonants, and as such they will facilitate a discussion of haplogroup B data. Turning now to macro-cultural relationships that facilitate a discussion of haplogroup B data, one of the defining characteristics of African cultural traditions is food production based either on pastoralism, hunting and gathering, or sedentary cereal agriculture. The herding of goats and sheep represent an important cultural relic of Afro-Asiatic languages in northern Africa. Cattle pastoralism is practiced among many Nilo-Saharan speaking populations in eastern Africa. Among the Hadza and Sandawe of Tanzania, the Khoisan of southern Africa, and the Pygmies of the central African rainforest, hunting and gathering stands as the traditional subsistence strategy. Finally, the origins and expansion of Niger-Congo speakers throughout sub-Saharan Africa evolved from the cultivation of millet and sorghum.

Within the Niger-Congo language family a discussion of haplogroup B data is also facilitated by a distinction between Bantoid and non-Bantoid languages. In order to discuss the term "Bantoid" it should be noted that unraveling the complex linguistic relationships within the Niger-Congo language family initially separates the approximately fifteen hundred languages of this family into three main branches: Mande, Kordofanian and Atlantic. According to *Ethnologue* (2016), the seventy-three languages of the Mande branch are concentrated in central western Africa. The same source places the twenty-three Kordofanian language in southern Sudan, and as such, they clearly occupy an "outlier" position within the Niger-Congo geographic distribution. The remaining and overwhelming majority of Niger-Congo languages (around fourteen hundred according to *Ethnologue*) fall within the Atlantic-Congo branch. These languages extend from Nigeria to South Africa. Nested deep within the numerous and complex sub-branches of Atlantic-Congo are about six-hundred languages classified by *Ethnologue* (2016) as "Bantoid." The reader, at this point is directed once again to the language map of Africa as provided in Supplementary Figure 2.2 and the distribution of Bantoid and non-Bantoid languages.

Section 3. The Macro-Linguistic and Macro-Cultural Distribution of B-M150 and B-M112.

Previously it was mentioned that B-M150 and B-M112 represent most of the haplogroup B diversity in Africa. Supplementary Tables 3.1 through 3.14 were prepared in order to report the available B-M150 and B-M112 data for the populations that speak languages that fall within the African macro-linguistic and macro-cultural paradigm as just described above in Section 2. Among the Bantoid populations, B-M150 is a clearly a significant marker (see Supplementary Table 3.8). About three quarters of the surveyed populations have the mutation. Moreover, the mutation extends along the entire geographic expanse of the Bantu expansion, such as among the Ngumba people of Cameroon and the Zulu of South Africa. The B-M112 mutation, on the other hand, stand as an insignificant marker among the Bantoid speaking populations as a whole (see Supplementary Table 3.1).

Among non-Bantoid populations, the B-M150 mutation appears in about a third of the surveyed populations, such as the Yoruba of Benin and the Fali of Cameroon (see Supplementary Table 3.9). According to the available data, the B-M150 mutation appears to have less significance among non-Bantoid populations as compared to Bantoid speaking populations. Nevertheless, the frequency of the mutation among some non-Bantoid populations points to the presence of the mutation in a western central Africa homeland prior to the Bantu expansion. These data agree with Scozzari et al. (2012) and their determination that the B-M150 mutation predates the Bantu expansion. Turning now to the B-M112 marker, just like among Bantoid populations, this mutation fails to attain a significant frequency among non-Bantoid speakers (see Supplementary Table 3.2).

The reader is now directed to Supplementary Table 3.3. As shown by the data, the B-M112 mutation stands as the signature marker of the so-called Pygmy populations. Among Baka populations in Gabon and Cameroon, the reported frequency is sixty percent or greater. A similar figure is attained among the Mbuti of the Congo region. A significant frequency of the B-M112 mutation is also found among the Aka of the Central African Republic and the Gyele of Cameroon. However, the B-M112 mutation is not found among two Pygmy populations, the Baganda of Uganda, and the Babinga of the Congo. Focusing now on Supplementary Table 3.10, the B-M150 mutation does not appear to be a significant marker among the Pygmies.

As shown by Supplementary Table 3.11, the B-M150 mutation does not appear from an overall perspective to be a significant marker for the Khoisan. However, the marker attains a perplexing frequency of seventy-nine percent among the ||Gana people of Botswana, and as such, presents a topic for future investigation. Perhaps, like the Damara of Namibia, the ||Gana were a Bantu group that switched to a Khoisan language (see Rocha and Fehn 2016). Turning now to Table 3.4, like the B-M150 mutation, based on an overall perspective, the B-M112 mutation has not attained a significant frequency among the Khoisan. Nevertheless, forty-seven percent of the Ju|’hoan people of Lesotho have the mutation.

The Hadza and the Sandawe people of Tanzania, as noted earlier, are counted among the African populations that speak an isolate language. Among both populations the B-M112 mutation attains a significant frequency, present in about half the Hadza and a third of the

Sandawe (see Supplementary Table 3.5). The B-M150 mutation, on the other hand, attains a frequency of ten percent among the Hadza according to one study, whereas two other studies failed to detect the mutation in the same population. Among the Sandawe, the B-M150 mutation is non-existent (see Supplementary Table 3.12).

With the possible exception of Cushitic speakers in Tanzania, the B-M112 mutation does not represent a significant marker for the Afro-Asiatic populations of Africa (see Supplementary Table 3.6). Turning now to B-M150, this mutation is virtually absent among these populations (see Supplementary Table 3.13). Surprisingly, the B-M150 mutation attains a significant frequency among some Nilo-Saharan populations: fifty-six percent of Alur (Congo region), twenty-two percent of Luo (Kenya), seventeen percent of the Ng'arkarimojong in Uganda. The same mutation is found in less than ten percent of Maasai (Kenya), Kanuri (Cameroon), and Sara (Chad). See Supplementary Table 3.14. The B-M112 mutation, on the other hand, fails to attain a significant frequency among the Nilo-Saharan populations of Africa (see Supplementary Table 3.7).

At this point it is important to mention several caveats about the data that was extrapolated from Supplementary Tables 3.1 through 3.14. First, the Nilo-Saharan data for B-M112 seems incomplete as Hassan et al (2008) only sequenced for B-M60 (the main haplogroup), and not the informative downstream variants, B-M150 and B-M112. Based on the results of B-M60 as reported by the study, B-M112 may well have a more significant frequency among the Nilo-Saharans. Secondly, Barbieri et al (2016) failed to report B-M112 data for each of the Khoisan groups that they tested. However, they aggregated the B-M112 data for the Khoisan groups as a whole and report an overall frequency of about twelve percent for these populations.

Thirdly, Barbieri et al. (2016) assert that the B-M150 mutation was already present in the Khoisan when the Bantus arrived in southern Africa about fifteen hundred years ago. The overall picture of B-M150 data, as presented in Supplementary Tables 3.8 through 3.14, seems to paint a different picture, confirming the opinion of other researchers, that B-M150 is a signature mutation of the Bantu expansion, and B-M112 mutation represents the genetic signature of hunter gatherer populations in Africa (e.g. Berniell-Lee 2009; Batini et al. 2011). Among the Khoisan, the most likely source of B-M150 mutations is probably geneflow from Bantu males who became part of the Khoisan groups.

A fourth caveat deals with the concept of ascertainment bias. The data reported in Supplementary Table 3.1 through 3.14 are extrapolated from populations for which, in many cases, a relatively small number of samples were collected for sequencing. As such, the reported mutation frequencies may not reflect the actual frequency found within the group. The scarcity of African data stands in sharp contrast to studies elsewhere in the world, especially Europe or East Asia, where hundreds of samples from a single population are sometimes sequenced in order to generate data. A final caveat stems from the fact that Africa possesses enormous cultural and linguistic diversity. Supplementary Tables 3.1 through 3.14 only capture a miniscule amount of this diversity. Hopefully the future will bring more data for more populations.

Section 4. The Bantu Expansion.

Obviously more genetic data from African populations is needed. Despite this handicap the available haplogroup B data nevertheless points to the B-M112 mutation as the genetic signature of hunter gatherer populations in Africa, and the B-M150 mutation stands as the genetic signature of the Bantu expansion through the central African rainforest and into southern Africa. As such both mutations help to assess male geneflow between Bantus and the hunter-gather populations of Africa. Thus haplogroups B-M150 and B-M112 help to identify factors that contributed to a massive language shift among the Pygmies, a topic discussed in this section.

It should be emphasized that Pygmies hunter-gatherers inhabited the central African rainforest long before the arrival of the farmers. After farming arrived in the rainforest, the Pygmies languages became extinct because they shifted to the languages of the farmers. An examination of the languages spoken by contemporary Pygmies reveals some Pygmy groups speak a Bantoid language as the result of contact with Bantu farmers. However, the Mbuti people speak Nilo-Saharan languages, which presumably came from an independent incursion into the central African rainforest by Nilo-Saharan-speakers from the Sudan. Moreover the Baka speak a non-Bantoid language from the Niger-Congo family that presumably came from an incursion into the rainforest by Ubangi-speaking farmers. This discussion of contemporary Pygmy languages raises a troubling issue due to the limited genetic data. An exploration of language shift among the Pygmies is limited primarily to interactions between this group and the Bantus. Again, as already noted, geneticists consider the B-M150 marker to be the genetic signature of the Bantu expansion.

Focusing now on the Bantus, the homeland of Niger-Congo languages encompasses an area of grassland that straddles the present-day border of southeastern Nigeria and northwestern Cameroon. As illustrated by supplemental maps provided by Grollemund et al. in their 2015 study, the northern boundary of the present-day central African rainforest is found in southern Cameroon and along the border separating the Central African Republic and the Democratic Republic of the Congo. As the result of climate change that occurred roughly four thousand years ago, savannah began to appear along the northern periphery of the rainforest. The appearance of savannah enticed people from the Niger-Congo homeland to move south according to Bostoen et al. (2015). As detailed in the same study, around 2.5 thousand years ago, savannah appeared in areas of the rainforest itself, which facilitated travel into the region, and eventually allowed the Bantus to travel through the region along the extensive network of rivers found in the area. Additionally, according to the study, Bantu farmers were able to utilize the areas of savannah within the forest to cultivate cereal crops. The primary African cereals cultivated by the Bantu farmers consisted of pearl millet, finger millet and sorghum (e.g. Crowther et al. 2017).

Serge Bahuchet in his 2012 report presents from an anthropological perspective several key points that help to explore the socio-linguistic history of Pygmies and farmers, including the Bantus. The term “Pygmy” has a racial component that reflects the short stature of the people forming approximately twenty different groups such as the Aka, Baka and Mbuti. They may have inhabited the rainforest for tens of thousands of years before the arrival of farmers. According to some oral traditions, when the Bantus and other farmers eventually penetrated the rainforest (about three thousand years ago), the Pygmies initially guided them through the “forest world.” Over time the farmers cultivated areas of the rainforest. Farmer groups then traded with nomadic or semi-nomadic Pygmy groups in order other to exploit a “common ecosystem.” One commodity of this exchange economy was Pygmy “brides.” Through marriage with Bantus, Pygmy women became part of many farming groups.

In his 2012 report Bahuchet examines the linguistic distance between contemporary Pygmy groups and their closest farming neighbors. This explains how the Pygmies shifted languages. In some case, the linguistic distance is close, which in turn indicates that both groups have lived alongside each other for a considerable period of time. Thus intense contact over a prolonged period of time seems to explain language shift among some Pygmy groups. However, in other cases the linguistic distance between contemporary farmer and Pygmy groups is large even though both groups live alongside each other. Furthermore, contemporary exchange between these neighboring Pygmy and farming groups is often facilitated by bilingualism rather than language shift on the part of the Pygmies. Bahuchet was able to identify linguistically close Pygmy and farmer groups that are now separated considerable geographic distance, sometimes several hundred kilometers. According to Bahuchet this indicates that sometime in the past both groups co-migrated, perhaps along a river, and then separated. Thus while the contact may have been intense, the duration of contact may have been relatively short.

As already noted, some of the Pygmies groups shifted to the language of the Bantu groups that they encountered. Often language shift is preceded by a period of intense contact between two groups where the social standing of one group is perceived as more prestigious than that of the other group. The Y-chromosome evidence suggests a more egalitarian relationship between Pygmies and Bantus. As reflected by Supplementary Table 3.1, the frequency of B-M112 among the Bantus fails to support a scenario in which a large number of Pygmy men had joined the various farming groups. Similarly, as reflected by Supplementary Table 3.10, the frequency of B-M150 among the Pygmies fails to support a scenario in which a large number of Bantu farmers had joined the Pygmy groups. Thus from an overall genetic perspective, male geneflow between Bantu farmers and Pygmies, as measured with the Y-chromosome mutations B-M150 and B-M112, appears to have been small. However, female geneflow between both groups has been measured with mitochondrial DNA and the data results from this marker are rather interesting. According to the available mitochondrial DNA data (Quintana-Murci et al. 2008) female geneflow from the Bantus to the Pygmies had not occurred. However, significant female geneflow between Pygmies and farmers had occurred. This conforms to the anthropological record, as presented by Bahuchet 2012, whereby one "commodity" of Bantu and Pygmy trade was Pygmy brides. Perhaps then, the concept of hypergamy helps to explain language shift among the Pygmies. Hypergamy describes situations where women marry men of higher socio-economic ranking. For the Bantus and Pygmies this explains why Pygmy women married Bantu men, whereas marriage between Bantu women and Pygmy men had not occurred. Taking this a step further, the Pygmies shifted to the language of their closest Bantu neighbors as farmer languages were considered more prestigious. This, in turn, invites further research that identifies factors or aspects of farming cultures that create prestige among hunter-gatherers. Such research has the potential of building more persuasive models of linguistic prehistory.

The suggestion of prestige motivated language shift among the Pygmies merely represents a working hypothesis, which in turn, is posited to encourage researcher to further explore agriculture as a vehicle for language shift, a very complex topic found not only in Africa, but elsewhere in the world. At this point some important caveats are in order. First and foremost, as suggested earlier, the amount of available genetic data, especially for Pygmies, are very limited. Additionally, the E-M180 mutation has also emerged as the genetic signature of the Bantu expansion (Paper 5.5. Hg. E. Section 7). However, the frequency of this mutation among the Pygmy groups remains very much a mystery. Berniell-Lee et al (2009) report twenty percent based on small sample sizes from three populations. Clearly, more data is needed. Nevertheless, E-M180 may reveal greater male Bantu to

Pygmy geneflow than what is suggested by the B-M150 mutation. This, in turn may suggest more intense contact between Bantus and Pygmies, an additional factor, in addition to prestige, that may have led to language shift.

Section 5. "Click Languages."

Click consonants are a very rare group of phonemes within the phonological inventory of language. However, these consonants stand as very productive speech sounds among the so-called Khoisan languages of southwestern Africa. Sandawe and Hadza, two of the isolate languages of Africa, also utilize click consonants. This raises an interesting question, whether Khoisan, Sandawe and Hadza share a common linguistic history given the fact that clicks are a rare phoneme. Alternatively, since the Hadza and Sandawe live in Tanzania and are separated by a distance of over two thousand kilometers from the Khoisan, clicks may have evolved independently in several African languages.

The B-M112 mutation has evolved into an informative haplogroup for exploring the history of African click languages because it is found both in Khoisan populations as well as among the Hadza and Sandawe. Knight et al. (2003) initially focused on B-M112 variation as found among the Hadza and mixed Khoisan samples. Based on their analysis of the B-M112 mutation and short tandem repeat (STR) variation found among the Khoisan and Hadza, the study suggested that both populations diverged about 120 thousand years. However, this estimate has a huge margin of error of plus or minus 40 thousand years, and therefore is not very accurate. Tishkoff et al (2007) broadened the examination of African click languages by including the Sandawe along with a mixed Khoisan sample and a sample from the Hadza. Based on analysis of the B-M112 mutation and short tandem repeat variation, they estimated that Khoisan, Sandawe and Hadza diverged from an ancestral population about 35 thousand years ago. With a margin of error of plus or minus four thousand years, their estimate is probably far more accurate than the one provided by Knight et al. (2003).

The significance of the studies by Knight et al. (2003) and Tishkoff et al. (2007) is that click consonance have significant time depth as indicated by the estimated the divergence of eastern African click-speaking populations and southern African click-speaking populations. Both studies reasoned that since click consonants are an especially rare phoneme, this speech sound had not evolved independently among the Hadza, Sandawe and Khoisan. Rather, click consonants stand as a relic of an ancient ancestral population from which all three populations descended. Both studies further suggest that click sounds initially evolved to aid hunters in their pursuit of game and later evolved into the earliest contrastive consonants. Such a scenario certainly fits the traditional hunter-gatherer subsistence strategy of all three groups.

Two important caveats are in order here. First, among the Khoisan groups, as reflected by Table 3.4, the B-M112 mutation only attains a significant frequency among the Ju|'hoan. According to the available data, the B-M112 frequency is low or non-existent in other so-called Khoisan groups. One possible explanation is that admixture with Bantus has diluted the frequency of haplogroups A and B in some Khoisan populations, whereas little admixture has occurred among the Ju|'hoan. Secondly, Güldemann and Stoneking published a report in 2008 that questions the finding presented by Tishkoff et al. (2007). They assert that the researchers made conclusions based on insufficient information and other factors may account for the clicks consonants found in African languages, such as independent innovation or language contact.

Support for the position taken by Güldemann and Stoneking (2008) stems from the fact that language contact has probably facilitated the introduction of clicks in some African languages. Xhosa, for example, is a Niger-Congo Bantoid language of South Africa that has click consonants as part of its phonemic inventory. In the case of the Xhosa people, Rocha and Fehn (2016) observed a high frequency of indigenous Khoisan mtDNA haplogroups. This indicates that a substantial number of Khoisan women have joined the Xhosa people through marriage with Xhosa men. Thus, in the case of Xhosa click consonants, language contact probably best explains their presence in the Xhosa language as clicks are not a regular feature of Niger-Congo languages. However, given the distance between Tanzania and South Africa, language contact seems to be a rather implausible explanation for the presence of clicks in the Sandawe, Hadza and Khoisan.

Moreover, the suggestion that clicks evolved independently among the Khoisan, Sandawe and Hadza also seems rather implausible. Contrastive click consonants are extremely rare and are only found in African languages. Some may cite Damin, a special register language found in Australia, as evidence of the potential of clicks to evolve independently. However, Damin is not classified as a language by *Ethnologue*, and even it were, clicks are still extremely rare. Thus given the choice of a common ancestral language, language contact, or independent evolution, the most likely scenario is that Sandawe, Hadza and Khoisan all share a common ancestral language that has roots extending deep into human prehistory when languages first evolved in Africa. Clicks, therefore, potentially stand as a relic of early language.

Section 6. Conclusions.

Like haplogroup A, haplogroup B-M60 also evolved and remained in Africa. Within the B-M60 main haplogroup the B-M150 and B-M112 mutations have emerged as especially informative mutations for understanding cultural and linguistic diversity on the African continent. For linguists, the B-M150 and B-M112 mutations are important components of a model of language contact induced shift among the Pygmies. The B-M112 mutation further points to clicks as potential first phonemes in the great tapestry of global language variation.

Chapter 4: Haplogroup D-M174.

Section 1. Out of Africa Theories.

The reader is invited to find haplogroup D at the bottom of Supplementary Figure 1.1 from Chapter 1. This diagram depicts the mutational steps from Y-Chromosome Adam to the main haplogroups. As shown by the figure, haplogroup D-M174 diverged from haplogroup DE-M145, which had earlier diverged from haplogroup DR-M168, the ancestral mutation of all the “out-of-Africa” haplogroups. Besides haplogroup D-M174, out-of-Africa haplogroups that evolved from DR-M168 include E-M96, C-M130, G-M201, H-M2713, I-M170, J-M304, L-M20, T-M184, KR-M526*, M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207. The term “out-of-Africa” implies, of course, that these haplogroups evolved outside of Africa and as such, they are distinct from haplogroups A and B-M60 which evolved and remained in Africa. Thus haplogroups A and B-M60 convey the story of human evolution in Africa (see Chapters 2 and 3), whereas the out-of-Africa haplogroups stand as the genetic artifacts of the human colonization of Eurasia, Australia, the Americas, and Oceania (see Chapters 4 through 17).

Traditionally, archaeological interpretation of human skeletal remains has served as the data source for developing models of the out-of-Africa exodus (e.g. Groucutt et al. 2015). More recent models, especially in the last twenty years, have also considered the genetic evidence as well as the paleoclimatological record. Oppenheimer in his 2012 paper defines several different questions that archaeologists, anthropologists, geneticists and geoscientists attempt to solve with their out-of-Africa models. Among these questions are the following: (1) the number of migrations out of Africa; (2) where human exited Africa; (3) when humans exited Africa; (4) why human exited Africa; and (5) the dispersal routes to East Asia and Europe. For the purpose of exploring language prehistory from a Y-chromosome perspective, a working out-of-Africa model will address these questions in order to provide a platform for exploring the evolution of language.

It should be emphasized that the question of “when” represents an especially important topic for linguists because the out-of-Africa migration potentially sets the minimum age of language evolution. Noble and Davidson in their 1991 paper dated the evolution of language at around 32 thousands year ago based on their interpretation of the archaeological record. While their opinion may or may not represent mainstream thinking in 1991, their paper suggests that researchers should now revisit the question of language evolution based on the accumulation of multi-disciplinary data over the last twenty-five years. Recent archaeological, paleoclimatological and genetic data, especially the Y-chromosome data for Australia (see Chapter 6: Section 3; Chapter 13: Section 6) may well push the evolution of language to a point much further back into the past than previous estimates derived from the archaeological record. Language may well have evolved before the human tribe left Africa.

Groucutt et al. in Table 1 of their 2015 paper provide an overview of several out-of-Africa models. The most striking difference between the various models is the question of when our human ancestors left Africa. Estimates range from 40 to 130 thousand years ago. Since the timing of the out-of Africa migration may define the potential evolution of language, this present chapter attempts to narrower the timeframe for this event. This inquiry starts by suggesting that a good out-of-Africa model should concede the following points:

- *Homo sapiens* evolved in Africa.

- The Y-chromosome data support a single out-of-Africa migration during Marine Isotope Stage 5 (between 71 and 130 thousand years ago).

- Admixture between humans and Neanderthals support human occupation of the Levant during Marine Isotope Stage 5 (between 71 and 130 thousand years ago).

- The paleoclimatological evidence supports an out-of-Africa migration into the Levant during Marine Isotope Stage 5 (between 71 and 130 thousand years ago).

- The fossil evidence supports an out-of Africa migration into the Levant during Marine Isotope Stage 5 (between 71 and 130 thousand years ago).

- Fossil evidence supports human occupation of the Levant during Marine Isotope Stage 4 (between 57 and 71 thousand years ago).

- The paleoclimatological evidence supports human occupation of the Levant during Marine Isotope Stage 4 (between 57 and 71 thousand years ago).

- Dating estimates for haplogroups D-M174, E-M96 and C-M130 support human occupation of the Levant during Marine Isotope Stage 4 (between 57 and 71 thousand years ago).

- The fossil record sets the human colonization of Europe, East Asia and Australia during Marine Isotope Stage 3 (between 29 and 57 thousand years ago).

- The paleoclimatological record sets the human colonization of Europe, East Asia and Australia during Marine Isotope Stage 3 (between 29 and 57 thousand years ago).

- The phylogeography of Y-chromosome variation and dating estimates support the human colonization of Europe, East Asia and Australia during Marine Isotope Stage 3 (between 29 and 57 thousand years ago).

At this point it should be noted that the term "Marine Isotope Stage" has become a common time unit used for discussing the out-of-Africa exodus because this event may well have been triggered by climatic change. Consequently, researchers often borrow the Marine Isotope Stage standard from the geoscientists to carry a discussion of the paleoclimatological, archaeological and genetic data for the out-of-Africa migration. Additionally, geological time (e.g. Pleistocene and Holocene) are also used. Finally, archeological time (e.g. Paleolithic and Neolithic) based on human tool artifacts, such as spear points or hand axes, are sometimes used to carry the discussion. The reader is now invited to examine Supplemental Table 4.1, which provides an overview of the time standards just described.

The following discussion of the Last Ice Age and the current Holocene epoch also provides necessary background information that facilitates a discussion of the role that climate change played in the out-of-Africa migration, especially within a Marine Isotope Stage paradigm. The term "ice age" denotes a period of glacial ice expansion across Northern Eurasia and lower sea levels worldwide because water is trapped within the ice. According to some researchers (e.g. Eldredge and Biek) glacial periods or ice ages are caused by variations of the earth's tilt and wobble in relation to its axis as well as temporary increases in distance between the sun and the earth. Thus tilt, wobble and orbit are stable during so-called "interglacial periods," such as the current Holocene epoch that began 12 thousand years ago. Conversely, they are unstable during glacial periods, including the Last Ice Age, which began about 130 thousand years ago and ended about 12 thousand years ago. Taking this a step further, variations in the earth's tilt, wobble and orbit not only produced sudden fluctuations in the advance and retreat of glacial ice during the Last Ice Age, but also periods of extreme precipitation or drought that appeared and disappeared within various regions of the world. This stands in sharp contrast to relatively stable climatic conditions in the current Holocene epoch because the earth's tilt, wobble and orbit are now stable. Consequently, the glaciers have steadily retreated in the last 12 years and the melting ice has caused sea levels to rise.

Having now discussed prehistoric time standards as well as glacial and interglacial periods, a discussion of the "out-of-Africa" migration now follows. The starting point of this migration is Marine Isotope Stage 5 which began 130 thousand years ago and ended 71 thousand years ago. The beginning of Marine Isotope Stage 5 coincides with the beginning of the Last Ice Age, which is significant as it signals the beginning of unstable weather conditions that potentially motivated the out-of Africa migration. Blome et al. in their 2012 paper provide environmental context for an out of Africa migration during Marine Isotope Stage 5. Their study is a synthesis of a tremendous amount of paleoclimatological and archaeological data from all of Africa which they divide into four different regions for comparison purposes: Southern Africa, Tropical Africa, East Africa, and North Africa. Figure 15 of their study illustrates the density of archeological sites within these four regions, from 30 to 150 thousand years ago, as well as periods of arid and humid conditions within this timeframe. From this figure, around 100 thousand years ago climate change would have placed an extraordinary amount of pressure on hunter-gatherer populations within Africa to migrate northwards in search of food. As shown by Figure 15, North Africa experienced humid conditions at this time, whereas the other regions of Africa were dry. Similarly, during Marine Isotope Stage 5 the Levant, like North Africa, also experienced humid conditions (see paleoclimatological data from Frumkin et al. 2011).

Based on the above discussion, climate data point to an out-of-Africa exit during Marine Isotope Stage 5 via North Africa and the Sinai, into the Levant. Fossil evidence also supports an out-of-Africa migration into the Levant during Marine Isotope Stage 5. Beginnings in the 1930s archaeologists have discovered several remains from Neanderthals and early modern humans from the Qafzeh and Skhul caves near the Sea of Galilee in Israel. The human remains are according to Oppenheimer (2012) between 90 and 120 thousand years old. However, it should be noted that Oppenheimer and other researchers (e.g. Mellars 2006) believe that the Qafzeh and Skhul represent an unsuccessful out-of-Africa migration based on their opinion that the exit occurred roughly 60 to 70 thousand years ago, which places the event within Marine Isotope Stage 3. However, the relatively recent paleoclimatological data from Blome et al (2012) and Frumkin et al. (2011), as just presented, make especially persuasive arguments that place the out-of-Africa exit roughly 100 thousand years ago during Marine Isotope Stage 5.

As noted by Oppenheimer in his 2012 paper, mtDNA and Y-chromosome data point to a single out-of-Africa migration. For the Y-chromosome, the genetic artifact of this migration is the DR-M168 mutation, which is the ancestral mutation for all the out-of-Africa haplogroups (see Supplementary Figure 1.1). (See, also, Underhill and Kivisild 2007). Dating estimates from Poznik et al. (2016) reflect that the DR-M168 mutation evolved about 91 thousand years ago. This estimate correlates well with an out-of-Africa migration during Marine Isotope Stage 5 as suggested by the fossil and paleoclimatological data.

Additional support for the Levant as the putative homeland of non-African Y-chromosome genetic diversity stems from studies that report genetic admixture between early modern humans and Neanderthals. Green et al. (2010) report that between one and four percent of human genetic inventory (or genome) contains Neanderthal DNA. Neanderthals were an archaic hominid that became extinct about 30,000 years ago. Apparently some mating (admixture) occurred between them and a small number of early human modern humans (e.g. Currat et al. 2011). Neanderthal DNA may have then boosted the immune system among early modern humans (Abi-Rached et al. 2011) and consequently was not eliminated from the human genome by recombination.

According to Green et al. (2010), Neanderthal DNA is only found in non-Africans. As such, they suggest that admixture between Neanderthals humans occurred in the Levant, a conclusion supported by the fossil record which places the Arabian Peninsula outside the known range of Neanderthals. Furthermore, Green et al (2010) suggest that human and Neanderthal admixture occurred before separation of the ancestral population of Eurasians and the Aboriginal Australians. This stems from the observations that all non-African populations have "statistically indistinguishable" amounts of Neanderthal DNA (Reich et al. 2011). From a Y-chromosome perspective, the DR-M168 mutation stands as the genetic signature of the ancestral population of Eurasians and the Aboriginal Australians. As mentioned in the previous paragraph, DR-M168 evolved in the Levant roughly 100 thousand years ago. Since DE-M145 and CR-P143 separated from DR-M168 about 65 thousand years ago, initial human admixture with Neanderthals probably occurred sometime during Marine Isotope Stage 5. Additionally admixture likely occurred in the Levant because, as just mentioned, the Levant and not the Arabian Peninsula is part of the known Neanderthal range.

The separation of DE-M145 and CR-P143 from DR-M168 (about 65 thousand years ago) represents the beginning of Y-chromosome diversification that occurred during Marine Isotope Stage 4, which started 71 thousand years ago and ended 57 thousand years ago. Almost immediately after CR-P143 evolved (65 thousand years ago), haplogroup C-M130 evolved from CR-P143. Later, about 62 thousand years ago, haplogroups D-M174 and E-M96 evolved from DE-M145 (see Supplementary Figure 4.1; and Supplementary Table 4.2 and the estimated Y-chromosome split times). This expansion of human Y-chromosome diversity during Marine Isotope Stage 4 signals an expansion of the human tribe outside of Africa that probably occurred in the Levant. Data from Frumkin et al. (2011) indicate that climatic conditions in the Levant during Marine Isotope Stage 4 were conducive for population growth. The Arabian Peninsula, another potential location where the human tribe could have dwelled, experienced drought-like conditions during this period (see Parton et al. 2015) and as such, the climate was not conducive for the Marine Isotope Stage 4 human expansions in this region.

Additional evidence for human expansion in the Levant during Marine Isotope Stage 4 comes from the fossil record. A partial modern human skull dated to at least 55 thousand years ago, the transition time from Marine Isotope Stage 4 to Marine Isotope Stage 3, was found at the Manot Cave in Israel. The skull has some Neanderthal features that are absent from the remains at Qafzeh and Skhul (see Hershkovitz et al 2015). This further suggests that the out-of-Africa human tribe lived in the Levant during Marine Isotope Stage 4 because, as noted earlier, the Arabian Peninsula is outside the known range of Neanderthals.

Marine Isotope Stage 3 began 57 thousand years ago and ended about 29 thousand years, close to the onset of the Last Glacial Maximum or LGM. During Marine Isotope Stage 3 the human tribe in the Levant separated. Some remained in the region, some migrated eastwards, and some migrated westwards. The eastward Marine Isotope Stage 3 migration resulted in the human settlement of South Asia, East Asia and Australia. Important fossil remains from this migration come from Lake Mungo in Australia, which date to at least 46 thousand years ago (see Bowler et al. 2003). Remains from the Ta Pa Ling Cave in northern Laos, as reported by Demeter et al. (2012), also have a similar date.

Most researcher favor a "southern dispersal" coastal route as the path taken by the human expansion into East Asia and Australia (e.g. Mellars 2006; Stoneking and Delfin 2010; Oppenheimer 2012). According Oppenheimer (2012), the migration followed the coastline along South and East Asia in order to harvest food from the sea. One question this scenario poses is how the human tribe reached the Gulf of Oman to begin their coastal migration. Traditionally researchers assumed that the out-of Africa migrations entered the Arabian Peninsula via the Red Sea and the narrow Gate of Tears that separates East Africa and Yemen. Then the migration followed the southern coast of the Arabian Peninsula and crossed over into South Asia at the Straits of Hormuz, another narrow crossing point (e.g. Mellars 2006; Oppenheimer 2012). One problem with this traditional scenario is that recent research, as detailed above, has placed the initial human out-of-Africa human expansion somewhere in the Levant. Additionally, the fossil and archaeological records fail to support human occupation of the southern Arabian Peninsula during Marine Isotope Stage 3 (Bailey et al. 2007). Perhaps a more parsimonious scenario is that humans were drawn towards the Black Sea and Caspian Sea, and later migrated alongside the Euphrates and Tigris Rivers to the Persian Gulf, and from this point, migrated along the coast of South Asia.

Pope and Terrell in their 2008 paper provide environmental context for a southern coastal migration to East Asia and Australia during Marine Isotope Stage 3. According to the paper the Indian sub-continent experienced cold and dry conditions during Marine Isotope Stage 4, which also produced wild fluctuations in sea level. Such conditions would not have sustained sufficient marine resources to feed a coastal migration. However, warmer weather and the monsoon rains returned about 50 thousand years ago during Marine Isotope Stage 3 and the sea level became stable. According to Pope and Terrell (2008), stable sea level produced abundant marine resources that fed the coastal migration to East Asia and beyond.

During Marine Isotope Stage 3 *Homo sapiens* also migrated westwards out of the Levant into Europe. Hoffecker, in his 2009 paper, reports the discovery of Aurignacian artifacts in Poland and Bulgaria dated to around 48 thousand year ago, which provide the earliest evidence for the human settlement of Europe. According to the paper, these artifacts resemble those found in the Middle East. The earliest fossil evidence comes from the discovery of human remains found at the Peştera cu Oase cave in Romania. They are dated to about 40 thousand years ago (Trinkaus et al. 2003, Fu et al. 2014: Supplementary Information 2).

Müller et al. (2011), as well as Hoffecker et al. (2009), assert that Marine Isotope Stage 3 climate changes around 48 thousand year ago facilitated the migration of *Homo sapiens* into Europe. According the Hoffecker et al. (2009) at this time a temporary retreat of the ice glaciers enabled humans from the Middle East to claim central European territory that had been previously abandoned by the Neanderthals. Such a scenario would support the idea that at the beginning of Marine Isotope Stage 3 *Homo sapiens* migrated from the Black Sea and Caspian Sea, and as such, the road to Europe and East Asia may not have started in the Arabian Peninsula as some researchers have believed.

As noted earlier, during Marine Isotope Stage 3 the human tribe in the Levant separated almost simultaneously, with some staying in the region and others migrating either eastwards or westwards. Turning now to the genetic evidence, during Marine Isotope Stage 3 the Y-chromosome further diversified in the Levant with the evolution of G-M201, J-M304, H-M2713, L-M20 and T-M184 in the region. As the human tribe migrated eastwards, haplogroups Q-M242 and R-M207 evolved in Northern Eurasia, and haplogroups N-M231, O-M175, M-P256 and S-B254 in eastern Eurasia Asia. The westward expansion towards Europe produced haplogroup I-M170. Thus, all the main haplogroups had evolved before the onset of Marine Isotope Stage 2 and the Last Glacial Maximum (see Supplementary Tables 1.1 and 4.2 for additional information).

Section 2. A Working Out-of-Africa Hypothesis.

The above presentation of paleoclimatological, genetic and archaeological evidence reflects that all three data sources converge to paint a picture of the human migration out of Africa into the Levant around 100 thousand years ago. During Marine Isotope Stages 5 and 4, the human tribe in the Levant expanded. During Marine Isotope Stage 3, improved climatic conditions in South Asia and Europe facilitated the human settlement of these regions. Accordingly, these data form a working out-of-Africa model and hypothesis that will be used to carry a discussion of global Y-chromosome variation, and with that, a model of prehistoric language expansion shaped by climate change, isolation, and agricultural technology. This working out-of-Africa hypothesis also points to the evolution of human language more than 100 thousand years in the past. The reasoning here follows the best of two possible scenarios, either the independent multiregional evolution of language, or the alternative and more plausible explanation, that the human tribe already had language when they left Africa.

An important caveat is now in order. The working out-of-Africa model is subject to change. For example, Liu et al., in their 2015 paper, report of the discovery of forty-seven human teeth that they confidently date between 80 and 120 thousand years ago. These remains were found at the Fuyan Cave in southern China. Arguably, this places the settlement of East Asia much further back in the past than the Marine Isotope Stage 3 scenario just adopted for this resource guide. Accordingly, we emphasize that the working out-of-Africa model now represents the one and only *successful* out-of-Africa migration. Again, data from three different disciplines independently converge to support this model. Additionally, Wei and Li, in their recent commentary (2017) that appeared in *Science Bulletin*, seem to agree. While the arrival of humans in East Africa 80 thousand or more years ago seems possible, one important question remains. Did they survive the massive Toba volcano explosion that occurred in Indonesia about 75 thousand years ago? Wei and Li (2017) suggest that the Fuyan Cave people became extinct as a result of the catastrophe.

Section 3. Overview of Haplogroup D-M174.

The DE-M145 mutation was one of the first Y-chromosome polymorphisms that were discovered (e.g. Hammer 1994). One of the distinguishing characteristics of this mutation and its downstream variants, haplogroups D-M174 and E-M96, is the presence of a unique Alu insertion polymorphism. This explains why the literature sometimes describes DE-M145, D-M174 and E-M96 as positive for the Y Alu Polymorphism (or YAP+).

At this point the reader is invited to review Supplementary Figure 4.1 which provides an overview of haplogroup D-M174 and its informative variants. As previously mentioned, D-M174 evolved about 62 thousand years ago in the Levant. Today this mutation is found almost exclusively in East Asian populations where it attains an overall frequency of around eleven percent (e.g. Zhong et al. 2011). The arrival of the D-M174 followed the southern dispersal route as outlined in the out-of-Africa model presented in Section 1 (above). (See, also, Shi et al. 2008; Stoneking and Delfin 2010). This being the case, the relatively low overall frequency of haplogroup D-M174 in current East Asian populations seems somewhat odd because it represents a genetic relic of the human settlement of this region around 50 thousand years ago. Researchers suggest that around eight thousand years ago Neolithic farmers were predominately haplogroup O-M175 while hunter-gatherers were predominately haplogroup D-M174 (Qi et al. 2013). Wang et al. in their 2013b paper suggest that a massive expansion of farmers during the Neolithic essentially shoved haplogroup D-M174 to the "periphery" of East Asia, an idea that follows the elevated frequency of this mutation among Japanese and Tibetans (see Sections 4 and 5 below). Nevertheless, elevated levels of haplogroup D-M174 also appear sporadically in small isolated Tai-Kadai speaking populations of East Asia, such the Lakkia and Zhuang people (Gan et al. 2008; Zhao et al. 2010**)**. Data provided by Cai et al. (2011) also reflects that haplogroup D appears sporadically in East Asia among populations speaking Austro-Asiatic and Hmong-Mien languages. Thus the genetic history of some East Asian populations still records the human settlement of this region despite the massive Neolithic expansion of haplogroup O-M175 (see Chapter 15 for additional details).

Section 4. The Settlement of Japan.

As mentioned earlier, Pope and Terrell in their 2008 paper trace the southern dispersal route as one that followed the seacoast of southern and (later) eastern Asia (see Section 1 above). According to the study the coastal migration reached Japan between 37 and 40 thousands years ago. At the time ice glaciers had trapped much of the global supply of water, which caused the sea level to drop as much as 130 meters below the present level. This facilitated the human colonization of Japan because this region was connected to mainland Asia. Japan later became a collection of islands after the Last Glacial Maximum and the resulting rise in global sea levels (e.g. Aikens and Akazawa 1996). For the purposes of this discussion, the southern dispersal migration is significant in that it contributed the D-M174 haplogroup that later became a genetic signature of the prehistoric hunter-gatherer Jomon culture (e.g. Hammer et al. 2006). The Jomon people then lived in isolation until the arrival of the Yayoi people about two thousand years ago (e.g. He, Yungang et al. 2012). See Ono (2002) and Chapter 15: Sections 14 and 15, for additional details.

Sato (2014) conducted a large-scale study of over two thousand Japanese males and found that thirty-two percent have the D-M174 mutation. Within the D-M174 main haplogroup, almost all the genetic variation among contemporary Japanese consists of the downstream D1b-M55 marker, which researchers consider a Japanese-specific mutation (e.g. He, Yungang et al. 2012). Besides D1b-M55, the C1a1-M8 mutation also stands as a genetic relic of the Jomon culture (see Chapter 6: Section 4, for additional details).

Section 5. Haplogroup D and Tibeto-Burman.

The Tibetan Plateau encompasses most of the historical region of Tibet. Here the average altitude is around 4,000 meters (13,000 feet) above sea level. According to a recent report (Zhang et al 2016) the inhabitants of this region possess a unique evolutionary adaptation that enables them to survive at such an extreme altitude. The same study further suggests, based on archaeological remains, that nomadic hunter-gatherers first inhabited the plateau around 30 thousand years ago on a seasonal basis. Then about seven thousand years ago farmers began to cultivate millet in the Middle Yellow River region. A thousand years later millet cultivation expanded westwards to the northeastern rim of Tibetan Plateau. However, according to Zhang et al. (2016) the interior of the region was left to the hunter-gatherers until around 3,600 years ago when farmers began to cultivate barley on the Tibetan Plateau, a crop that is more resistant to cold and dry climate of the region.

Qi et al. in their 2013 study report data from over two thousand Tibetan males. According to the study around fifty-four percent of the sequences belong to haplogroup D-M174 and thirty-three percent belong to O-M175. Haplogroup O-M175 represents a Neolithic component among the Tibetans, a genetic relic of the westward expansion of agriculture into the region from China (see Chapter 15: Section 4). Haplogroup D-M174, on the other hand, reflects a much earlier hunter-gatherer component, a relic of the human colonization of East Asia. As noted earlier in Section 4 (above) the D1b-M55 mutation is a unique Japanese specific variant of haplogroup D-M174. Similarly, Tibetans also have their own haplogroup D-M174 variant, the D1a-P99 mutation. According to Qi et al (2013) the D-P99 mutation evolved about 19 thousand years ago towards the end of the Last Ice Age. Among Tibetans the most common variant of D1a-P99 is the D1a-P47 mutation. Data from Qi et al. (2013) suggest that the D1a-P47 mutation evolved around 10 thousand years ago which corresponds roughly to evolution of agriculture in East Asia.

Ethnologue (2016) classifies Tibeto-Burman as one of the two main branches of the Sino-Tibetan language family, the other being Chinese. The significant frequency of haplogroup D-M174 among the Tibetans presents an interesting possibility that haplogroup D1a-P99 or its variants could be the genetic signature of Tibeto-Burman languages found in South Asia. However, data for India (Sahoo et al. 2006; Trivedi et al. 2008) and Bangladesh (Gazi et al. 2006) report a low frequency of haplogroup D among the Tibeto-Burman speaking populations outside of Tibet. Rather, the studies suggest that an expansion of haplogroup O-M175 carried Tibeto-Burman languages from its putative homeland in Middle Yellow River region of China into South Asia (see Chapter 15: Section 4).

Section 6. Haplogroup D and the Andaman Islands.

The color 'black' is *negro* in Spanish, and *negrito* is a diminutive form that has been utilized to describe several small isolated populations of people in Asia whose appearance resembles that of the African Pygmies. These so-called "Negritos" include the Jarawa and Onge of the Andaman Islands, the Semang of Malaysia, the Maniq of Thailand, the Aeta and Ati of the Philippines. Because of their unique African-like appearance, some researchers have taken an interest in the Negritos to determine if they are a relic population from the out-of-Africa expansion during Marine Isotope Stage 3.

The Andaman Islands are found in the Bay of Bengal, which is part of the Indian Ocean, and as such comprises part of the southern dispersal route followed by the eastward out-of-Africa expansion. A study from 2003 (Thangaraj et al.) tested twenty-three Onge samples and four Jarawa samples and found that all the samples were unspecified variants of haplogroup D-M174. In 2017 Mondal et al. reported additional data for the Onge and Jarawa based on whole genome sequencing. Among the Asian populations, the Jarawa and Onge are genetically closest to the Japanese based on close haplogroup D variation. Furthermore, the study reports that contemporary Japanese, Jarawa and Onge separated from a common ancestral population around 53 thousand years ago, which potentially places the Jarawa and Onge as a relic of Marine Isotope Stage 3 dispersals.

Section 7. Conclusions for Haplogroup D.

In order to discuss the Y-chromosome data, including that for haplogroup D-M174, considerable time was spent developing a working out-of-Africa hypothesis. Following this model, haplogroup D-M174 probably evolved in the Levant and arrived in East Asia during Marine Isotope Stage 3. Today the marker is found almost exclusively in East Asian populations, mostly among contemporary Tibetans and Japanese. Among both populations, this marker stands as an artifact of hunter-gatherer populations that roamed East Asia before the evolution of agriculture. Haplogroup D-M174 also stands as an important marker for understanding the prehistory of contemporary Jarawa and Onge populations on the Andaman Islands. Finally, Haplogroup D-M174 similarities between Andaman Islanders and the Japanese help to confirm that the human tribe followed the southern dispersal route when they migrated to East Asia about 50 thousand years ago.

Chapter 5: Haplogroup E-M96.

Section 1. Overview.

The reader is invited to review Supplementary Figure 1.1 from the first chapter. Both haplogroup E-M96 and haplogroup D-M174 diverged from DE-M145. According to Poznik et al. (2016: Supplementary Table 10), this occurred roughly 62 thousand years ago. Furthermore, as explained previously in Chapter 4: Section 3, the DE-M145 mutation and its downstream variants, D-M174 and E-M96, have a unique Alu insertion (YAP) polymorphism. However, despite the phylogenetic closeness, the phylogeographic distribution of E-M96 and D-M174 are very much different. As explained in the previous chapter (Chapter 4: Section 3), haplogroup D-M174 plays a rather modest role in representing the genetic diversity of East Asia. Moreover, within this region haplogroup D-M174 only represents a significant evolutionary marker for three populations: the Japanese, Tibetans, and Andaman Islanders. Haplogroup E-M96, on the other hand, represents a significant evolutionary marker for understanding the evolutionary history of populations in Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa. Additionally, haplogroup E-M96 represents almost all the Y-chromosome genetic diversity in Sub-Saharan Africa, where over ninety-two percent of men have a variant of this haplogroup (Luis et al. 2004).

Among the geneticists (e.g. Abu-Amero et al. 2009) most support the position that haplogroups E-M96 and D-M174 evolved outside of Africa in the Middle East. Haplogroup D-M174 then migrated to East Asia about 50 thousand years ago (see Chapter 4 for more details). Haplogroup E-M96, on the other hand, "back-migrated" to Africa by around 56 thousand years ago (Poznik et al. 2016). Interestingly, some argue (e.g. ISOGG 2017) that haplogroup E-M96 evolved in Africa because almost all of the sub-haplogroups of E-M96 evolved on the African continent. However, as suggested by Poznik et al. (2016), a more "parsimonious interpretation" of the data places the origins of E-M96 in the Middle East because otherwise haplogroups D-M174 and C-M130, as well the FR-M89 paragroup, would have been part of the out of African migration, which seems inconsistent with the genetic evidence and archaeological record.

On the African continent, about 50 thousand years ago, diversification of haplogroup E-M96 began with the evolution of the E1-P147 and E2-M75 mutations. Since then haplogroup E-M96 has undergone extensive diversification producing what seems to be an extremely complex arrangement of phylogenetic relationships (see, for example, the ISOGG website). Indeed, among the eighteen main haplogroups listed at the bottom of Supplementary Figure 1.1, Haplogroup E-M96 has arguably the most complex internal phylogenetic structure of mutational variants. In order to facilitate a discussion of linguistically significant E-M96 variants, our presentation of data for this haplogroup has been divided in six different "clusters" each with a color designation: orange, yellow, blue, red, green and purple (see Supplementary Figures 5.1 to 5.4).

The origins and expansion of languages in Africa seem to correlate well with the expansion of agriculture on this continent. The reader may recall that the herding of cattle in East Africa correlates well with Nilo-Saharan languages (see Chapter 2: Section 3). Similarly, the cultivation of sorghum and millet in West Central Africa carries the history of Niger-Congo languages (see Chapter 3: Section 4). In this present discussion of haplogroup E-M96 the reader now encounters another important language family on the African continent, languages classified as Afro-Asiatic. Arguably, Afro-Asiatic languages also co-expanded with agriculture like Nilo-Saharan and Niger-Congo. In the case of Afro-Asiatic, the origin of the agriculture expansion is centered in Southwest Asia (or the Middle East). This agricultural expansion involved the cultivation of crops such wheat and barley, as well as the herding of goats and sheep. Since several haplogroup E mutations record the co-expansion of Afro-Asiatic languages and agriculture from Southwest Asia into Africa, Sections 2 and 3 (below) provides necessary background information that facilitates a discussion of the haplogroup E-M96 data.

Section 2. Evolution of Agriculture in Southwest Asia.

The 2005 book *First Farmers: the Origins of Agricultural Societies* by Peter Bellwood provides an excellent resource for linguists who wish to explore the worldwide correlation between the origins of agriculture and the expansion of languages. In chapter three of the book (pp. 44-66) he explores the origins of agriculture in Southwest Asia, focusing on a region often identified in the literature as the "Fertile Crescent." This region encompasses parts of contemporary Egypt, Israel, Jordan, Lebanon, Turkey, Syria, Iraq and Iran. The transition to agriculture in the Fertile Crescent was facilitated by the domestication of cereals such as wheat and barley, and legumes such as chickpeas and lentils, from wild sources. Additionally, the agricultural transition in the Fertile Crescent involved the domestication of goats and sheep. The success of agriculture in Southwest Asia partly stems from improved climatic conditions following the Last Ice Age. Another factor that ensured the success of this transformation was the development of pottery.

Prior to the adoption of agriculture in Southwest Asia, and elsewhere in the world for that matter, the human tribe practiced hunter-gather techniques in order to survive. The evolution of agricultural in Southwest Asia generally follows a series of cultural transitions that began with the Natufians, followed by the Pre-Pottery Neolithic A and Pre-Pottery Neolithic B, and then finally the development of pottery itself. The Natufians stand as an important cultural transition because they were the last hunter-gatherers in the Middle East. According to Bellwood (2005) about 14.5 thousand years ago the Natufians appeared near the Sea of Galilee in what is now present-day Israel. Bar-Josef (1998) paints a picture of everyday Natufian life which centered on the hunting of gazelles and other animals. Moreover, and more significantly, he reports that they "practiced intensive and extensive harvesting of wild cereals" that grew abundantly in the region at the time. According to the description provided by Bellwood (2005) this abundant supply of food allowed the Natufians to construct semi-permanent settlements, something that is unusual for hunter-gatherers. These cultures are generally nomadic.

The Natufian thrived until about 13 thousand years ago when the Younger Dryas cold snap suddenly appeared. For a period of about seven hundred years global temperatures sank considerably. Weather conditions in Southwest Asia became cold and arid, and with that the abundant supply of wild cereals disappeared. Once again the Natufians became nomads and ultimately disappeared from the archaeological record (see Blockley and Pinhasi 2011**)**.

Then almost as suddenly as it began the Younger Dryas ended and warmer weather returned. This created ideal climatic conditions that produced, once again, what must have been a seemingly inexhaustible abundance of wild cereals (Bar-Yosef 1998; Bellwood 2005). Amid this abundance, for reasons not entirely clear, a significant human innovation occurred. People began to domesticate the wild cereals and legumes that their Natufian ancestors had previously gathered. The Pre-Pottery Neolithic A culture stands as the initial Southwest

Asian culture that embraced this new development. They and their descendants thrived and by around 10.5 thousand years ago large farming settlements appeared such as the one at Abu Hureyra in northern Syria. This development signaled the evolution of another cultural transition in the region, the Pre-Pottery-Neolithic B culture. One of the significant innovations that occurred during this period was the development of pastoralism, the herding of goats and sheep, which were once wild animals that people had managed to domesticate.

About nine thousand years ago the development of pottery ushered in a new cultural transition in Southwest Asia. This development allowed people to cook their food more efficiently and facilitated the storage of grain after harvesting. Around this time the climate in Southwest Asia also became more arid. According to Bellwood (2005) this change in climate was accompanied by deforestation that human settlements had brought as well as less productive soil due to over-farming. These conditions caused many people in Southwest Asia to abandon sedentary crop agriculture. Instead of cultivating crops, some turned to sheep and goat herding as a food source. By around 6.4 thousand years ago some of these Southwest Asian pastoralists herded their goats and sheep out of the region into Egypt (Kuper and Kroepelin 2006).

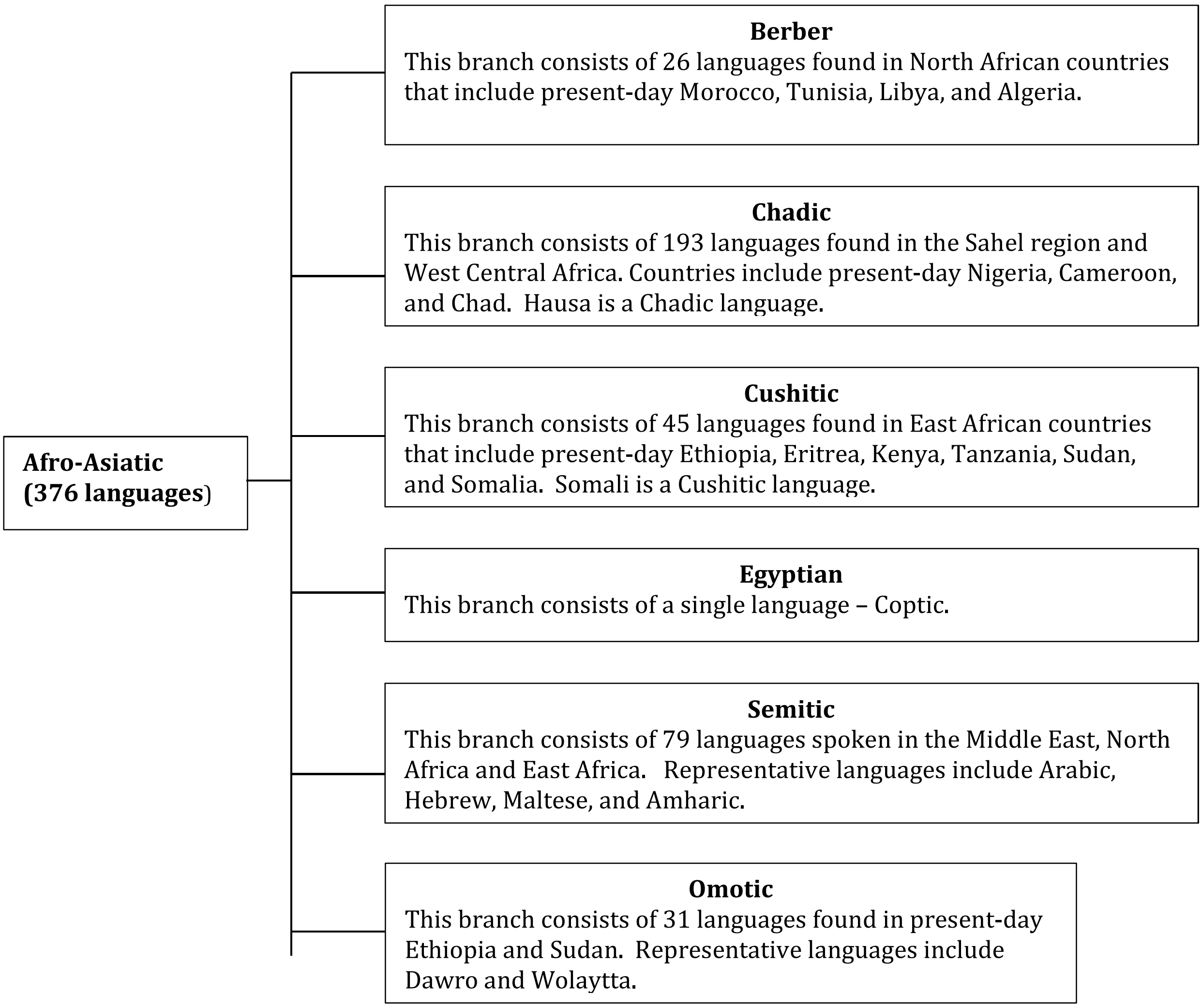
Section 3. Origins of Afro-Asiatic.

The Afro-Asiatic language family contains 376 languages (Ethnologue 2017). These languages are distributed throughout the Middle East, as well as in North Africa, East Africa, and West Central Africa. At this point the reader is directed to Supplementary Figure 2.1 from Chapter 2 which displays the distribution of Afro-Asiatic languages within Africa. Additionally, Figure 5.1 (below) lists the language branches of the Afro-Asiatic family and their contemporary geographic distribution. As shown by the figure, Afro-Asiatic is subdivided into six main branches: Egyptian, Semitic, Chadic, Cushitic, Omotic, and Berber. As inferred by the present-day distribution of these six main branches, Semitic evolved in Southwest Asia, whereas Chadic, Cushitic, Omotic, and Berber evolved in Africa. This scenario, of course, assumes that the current distribution of Arabic follows the historical spread of Islam.

Long-standing opinion among linguists (e.g. Ehret 2004) places the prehistoric origins of Afro-Asiatic languages somewhere in East Africa. This opinion follows the idea that most of the diversification within Afro-Asiatic occurred in Africa (e.g. Hetzron 2009). However, Bellwood (2005: 207-210), based on his interpretation of the archaeological data, suggests that Afro-Asiatic languages initially evolved in Southwest Asia and co-expanded out of this region with the spread of agriculture. Interestingly, linguistic data may also support this model of Afro-Asiatic origins. Using linguistic reconstruction, Militarev (2002) presents a proto-Afro-Asiatic lexicon of farming terminology. Based on the reconstructions, he suggests that the Natufians, agriculture and Afro-Asiatic co-evolved in Southwest Asia. Finally, another reason for identifying Southwest Asia as the putative homeland of Afro-Asiatic languages is the Y-chromosome data as presented below in Sections 4 and 5 (below).

Section 4. Green Cluster Mutations.

This paper employs color clustering as a tool for explaining the complex internal phylogeny of haplogroup E-M96. The reader is now invited to review Supplementary Figure 5.1 which provides an overview of the six E-M96 color clusters that will be used. As shown by the figure, the E1b1-P2 mutation unites the blue, red, green and purple clusters. The blue cluster represents downstream variants of the E1b1a-V38 mutation that evolved in West Central Africa. The green, red and purple clusters, on the other hand, first evolved in East Africa. Trombetta (2015) suggest that the diversification of E-P2 into these west and east variants occurred around 48 thousand years ago.



Afro-Asiatic (376 languages)

Berber
This branch consists of 26 languages found in North African countries that include present-day Morocco, Tunisia, Libya, and Algeria.

Chadic
This branch consists of 193 languages found in the Sahel region and West Central Africa. Countries include present-day Nigeria, Cameroon, and Chad. Hausa is a Chadic language.

Cushitic
This branch consists of 45 languages found in East African countries that include present-day Ethiopia, Eritrea, Kenya, Tanzania, Sudan, and Somalia. Somali is a Cushitic language.

Egyptian
This branch consists of a single language – Coptic.

Semitic
This branch consists of 79 languages spoken in the Middle East, North Africa and East Africa. Representative languages include Arabic, Hebrew, Maltese, and Amharic.

Omotic
This branch consists of 31 languages found in present-day Ethiopia and Sudan. Representative languages include Dawro and Wolaytta.

Figure 5.1. The Branches of Afro-Asiatic.
Data Source: Ethnologue 2017.

Once again the reader's attention is directed to Supplementary Figure 5.1. Note that the green, red and purple clusters evolved from E-M35, which evolved from E-M215, which evolved from E-P2. Trombetta et al. (2015) suggest that E-M35 arose in East Africa about 25 thousand years ago. This date is important as it provides time depth for the expansion of red and green cluster mutations out of East Africa into Egypt, and eventually into the Levant region of the Middle East. This second "out-of-Africa" migration probably followed the Nile River as it would have been an ideal corridor for human expansions (see Cruciani et al 2004; Luis et al. 2004; Cruciani et al. 2007; Cadenas et al. 2008). Note: the DR-M168 mutation represents the first out-of-Africa migration about 100 thousand years ago. See Section 1 (above).

Focusing now on the green cluster, Supplementary Figure 5.2 reflects that mutations within this cluster are variants of the E-Z827 haplogroup. One of the downstream variants is the E-PF1961 mutation. An interesting study from 2016 (Lazaridis et al.) was able to extract three ancient DNA samples from a Natufian archaeological site in Israel. As the reader may recall from Section 2 (above), the Natufians were the last hunter-gathers of Southwest Asia. Two of the samples belong to E-PF1961, which is an ancestral marker for E-M34.

The E-M34 mutation has a wide distribution, currently found in populations of Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa (see Supplementary Table 5.2). For linguists, the E-M34 mutation is significant as it represents a potential back-to-Africa marker that co-expanded from Southwest Asia into East Africa with goat and sheep herders about six thousand years ago. This Neolithic migration, as the reader may recall from Section 3 (above), appears to have carried Afro-Asiatic languages from Southwest Asia to East and North Africa. The ancient Natufian Y-chromosome data, as just presented, supports this position because the E-PF1961mutation is an upstream marker from E-M34. Thus, E-M34 probably evolved in the Middle East. Additionally, analysis of contemporary genetic data also supports Middle Eastern and agricultural origins of Afro-Asiatic languages in East Africa. Cruciani et al. (2004) suggest that E-M34 arose in the Levant based on their interpretation of the data and wider distribution of this marker outside of Africa. Additionally, Cadenas et al. (2008) date the E-M123 mutation in the Middle East to about 11 thousand years ago, which potentially places the evolution of the downstream E-M34 mutation during the Southwest Asian Neolithic.

Another significant green cluster mutation for linguists is E-M81. The position of the mutation within the green cluster phylogeny (see Supplementary Figure 5.2 and the contemporary distribution of the mutation (see Supplementary Table 5.1), suggest that E-M81 arose somewhere in Northwest Africa. This stems from the observations that the mutation exhibits a clinal frequency pattern across North Africa, with very low frequency among Egyptians, whereas the frequency climbs to around eighty percent among the Berbers of Morocco. In their 2004 analysis of the contemporary data Arredi et al. identify the E-M81mutation as a Neolithic marker. They suggest that goat and sheep pastoralism from Southwest Asia produced a "demic diffusion" of this mutation across North Africa and this expansion spread proto-Berber languages across the region. The term "demic diffusion" describes a scenario where a group adopts agriculture. This produces a sudden and rapid clinal population explosion across an uninhabited region because agriculture supports far more people per square kilometer than hunter-gathering food economies. Y-chromosome mutations frequently ride the coattails of such expansions. This explains why the E-M81 mutation has a low frequency in Egypt and a high frequency in Morocco.

The E-M81mutation consistently attains a high frequency among Berber populations (e.g. Bosch et al. 2001; Ennafaa et al. 2011; Fadhlaoui-Zid, et al. 2011; Trombetta et al. 2015). Accordingly this marker has become not only the genetic signature of the North African Neolithic but also the genetic signature of Berber languages. The Tuareg people of the Sahara desert also speak languages classified within Berber branch of the Afro-Asiatic language family. This suggests that these nomads are descendants of the North African Berbers. The genetic evidence, and more specifically, the elevated frequency of E-M81 among the Tuareg, supports this opinion (see Pereira et al. 2010; Ottoni et al. 2011).

A report detailing ancient DNA having the E-M81 mutation was recently posted by Fregel et al. (2017) on the bioRxiv website. The samples come from two remains found at the Ifri n'Amr o'Moussa archaeological site in Morocco. The remains are about five thousand year old and as such, they relics of the Neolithic in North Africa. These data provide additional support for contemporary DNA studies that equate the E-M81 marker as the genetic signature of the North African Neolithic.

The researcher Roger Blench (2014) posted a rather interesting paper on his website that presents an anthropological and linguistic perspective of the Berber people and language. According to the paper, a comparison of grammar suggests that Semitic is the closest Afro-Asiatic branch to Berber. Blench suggests that the Berber branch split from Afro-Asiatic language family around 6.5 thousand years ago. However, as he suggests, such a great time depth seem inconsistent with close linguistic similarities as found among the twenty-six contemporary Berber languages. Bench argues that a leveling of linguistic differences among the Berber languages occurred about two thousand years ago. This date is based on Neo-Punic and Latin lexical borrowings found in contemporary Berber languages. Blench suggests that the expansion of the Roman Empire into North Africa created a need for a *lingua franca* among the Berber. By this time the Berbers used camels and this brought an opportunity to trade with the Romans, especially along their southern frontier in North Africa, the so-called "limes." Thus, a *lingua franca* among the Berbers facilitated trade with the Romans. According to Bench, the adoption of a common trade language among the Berbers ultimately leveled linguistic diversity among this people. Blench further writes that the influence of the Berbers in North Africa later diminished after the spread of Islam throughout the region.

It should be noted that a discussion of the proto-Berber expansion across North Africa will continue in Chapter 10: Section 3 and the discussion of haplogroup J-M304. A variant of this main haplogroup, the J1-M267 mutation, which has origins in Southwest Asia, co-expanded with E-M81.

A final linguistically significant green cluster mutation is E-M293. Trombetta et al. (2015) estimate that this mutation evolved about 3.5 thousand years ago. In their 2008 study, Henn et al. suggest that the mutation evolved in Tanzania among the Datooga people. This population speaks a Nilo-Saharan language. However, as demonstrated by the data in Supplementary Table 5.3, the E-M293 mutation attains a significant frequency among several different populations, not only the Datooga. Furthermore, the languages spoken by these populations not only belong to the Nilo-Saharan language family, but also to Afro-Asiatic, Niger-Congo, and Khoe-Kwadi. Additionally, E-M293 is found among the Sandawe and Hadza, two populations that speak a language classified as an isolate.

Henn et al. (2008) suggest that the E-M293 mutation represents the genetic signature of an expansion of East African pastoralism and the herding of cattle, goats and sheep. According to the study, the migration began about two thousand years ago and covered territory that the Bantus migrated through about fifteen hundred years later. The migration of East African pastoralists apparently followed a corridor to South Africa that was free of the tsetse fly, a blood sucking insect capable of transmitting diseases which devastate livestock. This facilitated a demic diffusion scenario and the frequency of E-M293 expanded rapidly. Many of those with the mutation then joined hunter-gatherer societies. For linguists, what is particularly interesting about the E-M293 mutation is that agricultural expansions have the potential of producing a series of language shifts among the populations that is recorded by a genetic mutation. In the case of the southern expansion of East African pastoralism and E-

M293, the initial population may well have spoken a Nilo-Saharan language. As the migration expanded southwards, people shifted languages and adopted those that fall within the Afro-Asiatic, Niger-Congo, and Khoe-Kwadi language families, or the two isolates, Hadza and Sandawe.

Section 5. Red Cluster Mutations.

As noted previously in the discussion of green cluster mutations, haplogroup E-M35 evolved about 25 thousand years ago in East Africa. This date provides time depth for the co-migration of red and green cluster mutations out of East Africa into Egypt and later, the Middle East and Europe. Green cluster mutations are variants of the E-Z827 marker (Supplementary Figure 5.2) and red cluster mutation are variants of the E-M78 marker (see Supplementary Figure 5.3). Within the red cluster, five mutations represent potentially significant markers for linguists: E-V12, E-V32, E-V65, E-V13, and E-V22. Based on phylogenetic relationships as shown in Supplementary Figure 5.3, and frequency data shown in Supplementary Tables 5.4 through 5.8, the E-V12, E-V32, and E-V65 mutations probably evolved in Northeastern Africa, whereas E-V22 and E-V13 probably evolved in Southwest Asia.

As the reader may recall from Chapter 2: Section 3 and the discussion of haplogroup A-M13, about ten thousand years ago Holocene climate change transformed the Sahara desert into a savannah type ecosystem complete with rivers and lakes. Then about seven thousand years ago the rain stopped and the Sahara became once again a desert. As result of the so-called "desertification" of the Sahara, people either congregated along the Nile River in Egypt, or alternatively, moved with their herds of cattle, goats and sheep into the Sudan and East Africa (e.g. Kuper and Kröpelin 2006). Those that settled along the Nile eventually adopted sedentary agriculture and cultivated crops that came from Southwest Asia. The pastoralists, on the other hand, herded sheep and goats that came from Southwest Asia, and cattle that probably have an African origin (see Bellwood 2005: 97-103).

Hassan et al. (2008) suggest that E-V12 and E-V22 represent the genetic relics of the desertification of the Sahara. According to Cruciani et al. (2007), the E-V22 mutation evolved about ten years ago. Additionally, as noted earlier, the E-V22 mutation appears to have evolved in Southwest Asia. Based on this data, it appears that E-V22 may well represent a Neolithic back-to-Africa migration of farmers or pastoralists that spoke a proto-Afro-Asiatic language. Moreover, E-V22 may well have co-migrated into North Africa with the "green cluster" E-M34 mutation that has been described previously in Section 4.

E-V12, on the other hand, appears to have expanded after the arrival of agriculture in North Eastern Africa within a population that may well have spoken a Nilo-Saharan language. This scenario is supported by the earlier discussion that places the origins of the E-V12 mutation in North Africa. Additionally, *in-situ* origins and expansion of E-V12 is supported by dating estimates from Cruciani et al. (2007) who suggest that E-V12 evolved about fourteen thousand years ago. Since the E-V22 and E-V12 mutations currently attain a significant frequency in Nilo-Saharan and Afro-Asiatic speaking populations, it appears that language shift has occurred quite frequently in North and East Africa. In other words, the data paint a scenario suggesting that since prehistoric times Nilo-Saharan speaking populations have shifted to Afro-Asiatic, and Afro-Asiatic populations have shifted to Nilo-Saharan.

As shown by Supplementary Figure 5.3, the E-V32 marker is a downstream variant of the E-V12 mutation that was discussed in the previous paragraph. As noted earlier, E-V32 appears to have evolved in Northeast Africa. Dating estimates from Cruciani et al. (2007) suggest that this occurred about eight thousand years ago. According to the same study, E-V32 currently represents eighty-two percent of E-M78 (or red cluster) variation in East Africa. Frequency data from Supplementary Table 5.8, along with its estimated evolution date, suggest that the expansion of E-V32 in East Africa follows a demic diffusion model. In other words, the marker potentially represents an expansion of Nilo-Saharan cattle herders, or alternatively, an expansion of Afro-Asiatic speaking pastoralists, or alternatively both, from Egypt to East Africa.

Focusing now on the E-V65 mutation, very little information for this marker is currently available. As noted earlier, E-V65 probably evolved in North Africa. Dating estimates from Cruciani et al. (2007) suggest that this occurred about four thousand years ago. Data from Supplementary Table 5.7 indicate that E-V65 attains a significant frequency among Arab populations in North Africa, whereas the frequency among Berber populations is low. This mysterious variation in frequency numbers seems to be a topic worthy of additional research.

The E-V13 mutation is the only haplogroup E-M96 variant that attains a significant frequency in Europe. As shown by Supplementary Table 5.4, E-V13 attains a significant frequency among the populations of the Balkans and in Greece. More moderate frequencies are observed elsewhere in Europe, such as among the Italians and the Hungarians. Most studies suggest that E-V13 entered Europe during the Mesolithic (Battaglia et al. 2009; Regueiro et al. 2012; Karachanak et al. 2013). Regueiro et al. (2012) in their study of Serbs estimate the presence of E-V13 in the Balkans by 12 thousand years ago. Thus, the arrival of E-V13 may follow the disintegration of the Natufian culture during the Younger Dryas (see discussion in Section 2). It should be noted that a recent study (Trombetta et al. 2015) suggests that E-V13 evolved around eight thousand years ago. Here, researchers favor a Neolithic or latter arrival of the mutation in Europe. Regardless, the presence of E-V13 in the Balkans, either during the Mesolithic or Neolithic, raises an interesting question: Were proto-Afro-Asiatic languages part of the linguistic inventory of prehistoric Europe?

From the frequency data tables provided for this present chapter the reader may notice that in addition to E-V13, other haplogroup E mutations also appear among the populations of Mediterranean Europe and Iberia. It should be emphasized, once again, that only E-V13 attains a significant frequency in Europe. Moreover, E-V13 has a clear prehistoric presence on this continent. Turning now to appearance of the green cluster E-M81 mutation in Iberia, the literature almost always treats this as a genetic relic of the Islamic (or Umayyad) conquest of the peninsula in the year 711. While this mutation may attain a significant frequency among a few isolated populations in Spain and Portugal, it should be noted that the overall frequency of E-M81 in Iberia is otherwise low (e.g. Regueiro et al. 2015). As such the Islamic conquest of Iberia added little to the gene pool in contemporary Spain and Portugal. For haplogroup E-M96 variants other than E-V13 and E-M81, a pattern surfaces whereby these mutations are generally found at a low frequency among populations residing on the Mediterranean islands of Europe or along the European Mediterranean coast. Here, historical trade and cultural exchange between North Africa and Europe may well provide an explanation (e.g. Cruciani et al. 2007).

Section 6. The Purple Cluster E-V6 Mutation.

The only mutation within the purple cluster is E-V6 (see Supplementary Figure 5.1). Very little information is available about this marker and future research in this area might be fruitful. Cruciani et al. (2004) report that this mutation is found in nine percent of Ethiopians. A more recent study (Trombetta et al. 2015) suggests that the mutation attains a significant frequency among several East African populations speaking languages that belong to either the Afro-Asiatic or Nilo-Saharan language families (see Supplementary Table 5.9). The same study estimates that the mutation evolved around twelve thousand years ago. However, their phylogenetic placement of the mutation within the haplogroup E-M96 hierarchy is substantially different than that of the International Society for Genetic Genealogy (ISOGG). According to ISOGG (2017), the E-V6 mutation branches directly from E-M35, whereas Trombetta et al. place the mutation much further downstream within E-Z827. Thus based on the ISOGG phylogeny, E-V6 is potentially much older than the estimate provided by Trombetta et al. (2015). Taking this a step further, the E-V6 mutation potentially represents an E-M35 variant that remained in East Africa at a time when E-M78 and E-Z827 left the region. As such, E-V6 may stand as an ancient genetic relic of pre-agricultural East Africa.

Section 7. The Blue Cluster.

In Chapter 2: Section 3 and the discussion of haplogroup A-M13, a connection was drawn between East African cattle pastoralism and the linguistic prehistory of Nilo-Saharan languages. Sections 2 and 3 (above) also present a discussion of the evolution of pastoralism in Southwest Asia and the linguistic prehistory of Afro-Asiatic languages. Moreover, Chapter 3: Section 4 discussed the linguistic prehistory of Niger-Congo languages, the evolution of agriculture in West Central Africa, and the B-M150 mutation. Now the evolution of agriculture in West Central Africa continues with a discussion of blue cluster mutations, which are variants of E-V38.

The reader is invited, once again, to review Supplementary Figure 5.1. As shown by the figure, and as explained earlier, the E1b1-P2 mutation unites the blue, red, green and purple clusters. The blue cluster represents downstream variants of the E1b1a-V38 mutation that evolved in West Central Africa. The green, red and purple clusters, on the other hand, evolved in East Africa. Trombetta (2015) suggest that the diversification of E-P2 into these west and east variants occurred around 48 thousand years ago. Focusing now on Supplementary Figure 5.4, the reader finds linguistically significant variants of the E-V38 "blue cluster" mutation.

The E-M2 mutation represents a downstream variant of E-V38. Data from Poznik et al. (2016) suggest that E-M2 diverged from E-V38 about 40 thousand years ago. Within the E-M2 phylogeny (see Supplementary Figure 5.4), the E-U174 and E-U175 mutations have been identified as especially strong genetic signatures of the Bantu expansion from the Niger-Congo language family homeland in West Central Africa (e.g. Filippo et al. 2011; Montano et al. 2011; Barbieri et al. 2012; Rowold J. et al. 2016). This is based on frequency data for both mutations. For example, in their 2016 study of Bantu populations in Mozambique, Rowold et al. found that twenty-five percent of the samples belong to E-U174 and thirty-seven percent belong to E-U175. Besides the frequency data, dating estimates taken from West Central African populations also identify E-U174 and E-U175 as the genetic signature of the Bantu expansion. According to Filippo et al. (2011), the E-U175 mutation evolved in West Central Africa about five thousand year ago, and E-U174 evolved in the same region about four thousand years ago. These dating estimates agree with the timeframe for the Bantu expansion as taken from the archaeological record (see discussion in Chapter 3: Section 4).

In their survey of populations in West Central Africa, Filippo et al. (2011) and Barbieri et al. 2012 found that genetic diversity in Mande speakers and non-Bantoid Atlantic-Congo speakers to be older than in the Bantu populations. Mande and non-Bantoid Atlantic-Congo populations tend to have the orange cluster E-M33 mutation as well as undefined older mutations within E-M2 (see, also, Supplementary Figure 5.1; Supplementary Table 5.10; and Section 8 of present chapter). The Bantus, on the other hand, tend to have, almost exclusively, the E-U174 and E-U175 variants. This observation stands in general agreement with the linguistic evidence that places Mande and non-Bantoid Atlantic-Congo closer towards a theoretical Niger-Congo proto-language.

Filippo et al. (2011) also suggest that E-U174 and E-U175 variation found among Pygmy groups may well be undefined older variants of both mutations, and as such, this questions the extent of Bantu and Pygmy admixture. Perhaps the undefined older variants came from non-Bantoid populations before the Bantu expansion. Thus, additional resolution of E-M2 and its downstream variants is needed in order to further clarify the genetic history of the Bantus, other Niger-Congo speaking populations, and the Pygmies.

Section 8. The Orange Cluster E-M33 Mutation.

According to data from Poznik et al. (2016), E1-P147 diverged from the E-M96 main haplogroup about fifty thousand years ago. Shortly thereafter, about forty-eight thousand years ago, E-M33 diverged from E1-P147. These dating estimates, along with its position within the E-M96 main haplogroup phylogeny (see Supplementary Figure 5.1), reflect that the E-M33 orange mutation evolved shortly after the initial back-to-Africa migration, which occurred around 60 thousand years ago. As such, E-M33 represents a comparatively ancient mutation that traces its origins close to the initial diversification of E-M96 variation in Africa. As shown by Supplementary Table 5.10, the geographic distribution of E-M33 populations is rather interesting as these populations are found in the Sahel region of Africa, an area that stands as a transition region between the southern border of the Sahara desert and Sub-Saharan Africa. Moreover, this region represents the putative homeland of Nilo-Saharan languages (see discussion in Chapter 2: Section 3).

Within the Sahel region, Supplementary Table 5.10 reflects that E-M33 attains a moderate frequency among Nilo-Saharan, Niger-Congo and Afro-Asiatic populations. As noted previously in Section 7, the presence of E-M33 among Mande speakers and non-Bantoid Atlantic-Congo speakers confirms what the linguistic evidence suggests, that they are the ancestral populations of the Bantu. Moreover, the E-M33 data supports the idea that Niger-Congo and Nilo-Saharan languages may well share a common linguistic ancestor (see discussion in Chapter 2: Section 3). Turning now to Afro-Asiatic, the presence of the E-M33 mutation among Chadic speaking populations, such as the Kotoko and Masa of Cameroon (see Bučková et al 2013), and the Hausa of Sudan (see Hassan et al. 2008), might also be significant in that Chadic populations also have a significant frequency of the R1b-V88 mutation (see Paper 5.17, Section 7, and the discussion of haplogroup R1b-V88). This is significant because the discovery of a variant of the R1b-M343 mutation in the Sahel was unexpected and is still difficult to explain (e.g. Cruciani et al. 2010). R1b-M343 represents, otherwise, the genetic signature of West Eurasian populations.

Section 9. The Yellow Cluster E-M41 Mutation.

The reader is directed to Supplementary Figure 5.1 and the E-M41 “yellow cluster” mutation. Very little is known about this mutation, including when it diverged from E2-M75. Most of the data comes from a 2010 study (Gomes et al.) and one hundred and eighteen samples taken from three different populations in Uganda: the Dodoth, Jie and Karimojong. These groups speak Ng'arkarimojong, a Nilo-Saharan language. Overall, the E-M41 mutation attains a modest frequency of eleven percent. In their 2005 study Wood et al. report that this mutation attains a frequency of sixty-seven percent among the Alur people of the Democratic Republic of the Congo, a population that also speaks a Nilo-Saharan language. However, the sample size was very small (nine men) and ascertainment bias may well have skewed the actual frequency. The only other African population in which E-M41 attains a significant frequency is the Hema of the Democratic Republic of the Congo where the mutation is reported in thirty-nine percent of the men (Wood et al. 2005). This population speaks a Niger-Congo Bantoid language. Elsewhere in Africa (e.g. Luis et al. 2004), the M41 mutation attains a very small frequency.

Gomes et al. (2010) suggest that the E-M41 mutation represents a potential marker for understanding the genetic history of Nilo-Saharan speaking populations in East Africa. Indeed, the data suggest that the genetic relics of pre-agricultural Nilo-Saharan speaking populations in Africa are the E-M41 “yellow cluster” mutation, the E-M33 “orange cluster” mutation (see Section 8), and the E-V12 “red cluster” mutation (see Section 5).

Section 10. Conclusions.

Within the E-M96 main haplogroup, thirteen different variants stand as especially informative mutations for deciphering the correlation between linguistic and genetic diversity: E-M34, E-M81, E-M293, E-V12, E-V32, E-V65, E-V13, E-V22, E-V6, E-U174, E-U175, E-M33 and E-M41. The mysterious presence of E-V65 in North African Arabs requires additional research. The blue cluster E-U174 and E-U175 mutations carry the Bantu expansion southwards from West Central Africa. Proto-Berber and the green cluster E-M81 mutation co-expanded across North-Africa. The red cluster E-V22 and green cluster E-M34 variants represent Afro-Asiatic agriculturalist that entered North and East Africa during the Neolithic. Red cluster E-V12, orange cluster E-M33, and yellow cluster E-M41 are genetic relics of pre-agricultural Nilo-Saharan populations. E-V32 from the red cluster represents the expansion of Nilo-Saharan and/or Afro-Asiatic pastoralists into East Africa. Pastoralism later expanded from this region with the green cluster E-M293 mutation. This expansion triggered a series of language shifts among the herders who joined hunter-gatherer groups. The lone survivor of pre-agricultural East Africa appears to be the purple cluster E-V6 mutation. Finally, the red cluster E-V13 raises the possibility that some prehistoric Europeans may have spoken a proto-Afro-Asiatic language.

Chapter 6: Haplogroup C-M130.

Section 1. Overview of Haplogroup C-M130.

As discussed in Chapter 4: Section 1, the DR-M168 mutation represents the exodus of modern humans from Africa into the Levant roughly 100 thousand years ago. In this region, roughly 65 thousand years ago, three main haplogroups evolved from the M168 mutation: D-M174, E-M96, and C-M130. As discussed in the previously in Chapter 5, E-M96 back-migrated to Africa. However, C-M130 co-migrated out of the Levant with D-M174 (see Chapter 4) about 50 thousand years ago during Marine Isotope Stage 3.

Focusing now on the internal phylogenetic hierarchy of the C-M130 main haplogroup, it should be noted that two different nomenclature standards are currently being used to report C-M130 data. Some studies (e.g. Wei et al. 2017b) continue to use the phylogenetic tree as presented by Karafet et al. (2008). Other studies (e.g. Huang et al. 2017) utilize the current nomenclature standard that has been adopted by International Society for Genetic Genealogy (ISOGG). We utilize ISOGG (2017) and as such it should be noted that the current ISOGG perspective is far different than that presented by Karafet et al. (2008). Karafet et al. (2008) arranged the internal C-M130 phylogeny into five main branches: C1-M8, C2-M38, C3-M217, C4-M347, and C5-M356. However, ISOGG (2017) divides C-M130 into two main branches, C1-F3393 and C2-M217. For the sake of simplicity, the ISOGG perspective can be shortened and viewed as either *C1* or *C2*. Interestingly, the ISOGG (2017) perspective closely follows the archaeological record whereby *C2* expanded from a refugium in south central Siberia towards the beginning of the Holocene. *C1*, on the other hand, represents earlier human migrations during Marine Isotope State 3. These expansions signal the initial human settlement of India, Island Southeast Asia, Australia, Japan and Europe.

Section 2. Overview of *C1* Mutations.

At this point the reader is invited to review Supplementary Figures 6.1 and 6.2. Both focus on downstream variation within the C1-F3393 mutation. According to Poznik et al. (2016: Supplementary Table 10), C1-F3393 separated from C-M130 about 45 thousand years ago. Based on the genetic, archaeological and climatological data (Sections 3 and 4 below), it would appear that this may have occurred in northern India. Turning now to Supplementary Figure 6.1, this chart focuses on C1b-F1370 mutations that expanded along a southern dispersal route during Marine Isotope Stage 3. The same route also carried haplogroup D-M174 variants to East Asia (see Chapter 4). Supplementary Figure 6.2, on the other hand, focuses on C1b-F1370 mutations that expanded along a northern dispersal route during the same period.

Section 3. Eastward Expansion of C1b-F1370 Mutation across Eurasia via a Southern Route.

3.1. Overview.

Many researcher favor the human colonization of East Asia during Marine Isotope Stage 3 (roughly 50 thousand years ago) via a single southern route, the so-called southern dispersal hypothesis (see Chapter 4: Section 1). Turning now to Supplementary Figure 6.1, the reader will notice the use of color clustering to highlight particularly important mutations within the C1b-F1370 branch. The C1b-M356 "gold cluster" mutation represents the initial human colonization of South Asia and is only found in India, Nepal and Pakistan. The C1b-M38 "green cluster" mutation is confined to Island Southeast Asia and Oceania. This mutation stands as genetic relic of the human colonization of New Guinea and Indonesia. The human colonization of Australia is represented by the C1b-M347 "blue cluster" mutation, which is only found on this continent.

3.2. C1b-M356 Gold Cluster Mutation and Pleistocene Colonization of India.

As noted earlier, the C-M356 mutation is restricted to India, Nepal and Pakistan (see, also, Sengupta et al. 2006). Population studies reporting frequency data for C1-M356 indicate that this mutation attains a small frequency within the region, somewhere around five percent or even less. Despite the low frequency numbers, C1-M356 is still a significant mutation as it represents the genetic relic of the founding population of India (e.g. Sengupta et al. 2006; Arunkumar et al. 2012; Khurana et al. 2014). This agrees with the dating estimate provided by Poznik et al. (2016: Supplementary Table 10), who report that the C1-M356 mutation evolved roughly 44 thousand years ago.

Contemporary India has a huge population with almost 1.3 billion people (*CIA World Factbook* 2017). The traditional social hierarchy consists of either castes or tribes. Together with a large population and complex social structure, one finds incredible linguistic diversity. *Ethnologue* (2017) lists 462 languages for India. Almost all these languages fall within one of the four language families: Dravidian, Indo-European, Austro-Asiatic or Sino-Tibetan. Thus for linguists the C1-M356 mutation represents a starting point for gaining an understanding of the correlation between genetic and linguistic diversity in India. This entails the identification of mutations that are "Paleolithic" like the C1-M356 marker, and mutations that represent more recent migrations during the Mesolithic or Neolithic. Accordingly, the discussion of linguistic diversity on the Indian subcontinent continues in Chapter 8 with the presentation of haplogroup H-M2713.

3.3. C1b-B477 and the Colonization of Sunda and Sahul.

As noted above, variants of the C1b-F1370 mutation expanded into South and East Asia roughly 50 thousand years ago during Marine Isotope Stage 3. As shown by Supplementary Figure 6.1, C1b-F1370 splits into C1b-K281 and C1b-B477. Downstream from C1b-K281 is the C1b-M356 mutation, which represents, as discussed above, the human colonization of India. Focusing now on another C1b-F1370 variant, the C1b-B477 mutation, we find a genetic artifact of the human colonization of Australia and Island Southeast Asia. It should be noted that we define Island Southeast Asia as Indonesia and Papua New Guinea.

According to Karmin et al (2015: Table S7 and Figure S21), C1-B477 evolved about 49 thousand years ago. Downstream from C1b-B477, C1b-M38 is an informative mutation that evolved in Island Southeast Asia. C1b-M347, another informative mutation downstream from C1b-B477, evolved in Australia. In order to understand the human colonization of Island Southeast Asia and Australia, it should be noted that during the Last Ice Age sea levels were considerably lower than present day levels (e.g. Clark et al. 2009). As a result of lower sea levels, a large landmass called "Sunda" connected the present-day Malaysian Peninsula and many of the contemporary Indonesian Islands, including Sumatra, Java, Borneo, and Bali. At the same time the Sahul landmass connected Papua New Guinea and Australia (see

Supplementary Figure 6.3 for additional details). Since the distance between Sunda and Sahul may have been as short as ninety kilometers, a water crossing between both landmasses seems quite feasible even with primitive watercraft (Allen and O'Connell 2008). Additionally, fossil remains support the initial human colonization Sunda and Sahul. Lake Mungo man in Australia is dated to at least 46 thousand years ago (see Bowler et al. 2003). The so-called "Deep Skull" at the Niah Cave on the Indonesian Island of Borneo is at least 35 thousand years old (Barker et al. 2007).

3.4. C1b-M38 and Colonization of Island Southeast Asia and Oceania.

C1-M38 is found almost exclusively in Island South East Asia (e.g. Mona et al. 2007; Mona et al. 2009; Karafet et al. 2010; Tumonggor et al. 2014). Karmin et al. (2015) suggest that C1-M38 evolved about 24 thousand years ago. Based on their interpretation of the data, Mona et al. (2007) suggest that C1-M38 evolved in the northwestern part of New Guinea, and the mutation expanded both to the eastern part of the island (Papua New Guinea) as well as westward to Indonesia. For linguists, the C1-M38 mutation represents an indigenous component within the linguistic tapestry of Island Southeast Asia, and as such, stands as an informative marker for deciphering the evolution of the so-called "Papuan" macro-family of languages.

Besides Papuan languages, downstream variants of the C1-M38 mutation are also important mutations for decipher the evolution and expansion of the Austronesian language family. This language family has both a Papuan/Island Southeast Asian component and an East Asian/Taiwanese component. Focusing now on the East Asian component, both the linguistic and archaeological evidence suggest that this language family originated among the aboriginal populations on the island of Taiwan (e.g. Diamond 2000). Beginning about six thousand years ago, Austronesian-speaking populations expanded southwards from Taiwan into the Philippines (see Chapter 15: Sections 6 and 7). Another expansion then carried Austronesian into Papua New Guinea and Indonesian. From Papua New Guinea, a final expansion carried Austronesian eastwards across a vast ten-thousand-kilometer expanse of ocean. We define this vast expanse of ocean as Oceana, the numerous Pacific Ocean islands that extend from the Solomon Islands to Easter Island (Rapa Nui).

Similar to the linguistic and archaeological data, Austronesian-speaking populations have a Papuan component from Island Southeast Asia and an East Asian genetic component from Taiwan. The Taiwanese genetic contribution will be discussed in Chapter 15: Section 6 and 7. Focusing now on the Papuan genetic component, about 12 thousand years ago the C1-M208 mutation, which is a downstream variant of C1-M38, evolved in the highlands of western New Guinea (Delfin et al. 2012; Karmin et al. 2015). C1-M208 is rarely found in Indonesia (e.g. Mona et al. 2009; Karafet et al. 2010), which suggests a minimal westward expansion of the mutation. Rather the C1-M208 mutation exhibits an increasing frequency cline from New Guinea across Oceania (Mirabel et al. 2012).

When the Austronesians reached Papua New Guinea roughly three thousand years ago (see Chapter 15: Section 7) admixture occurred between this group and the Papuan who had lived on the Island for over 40 thousand years. Then roughly two thousand years ago, a subset of this New Guinean admixed population expanded eastwards across Oceania. The C1-M208 mutation represents the Papuan component of the admixed population that colonized this region.

As C1-M208 moved across Oceania, a downstream variant of this marker, the C1-P33 mutation, evolved roughly 4.5 thousand years ago, possibly in the Tongan archipelago (Cox et al. 2007). As seafarers moved further east across Oceania, the C1-P33 mutation eventually became the only genetic mutation among the populations that colonized the region, which suggest the effects of founder effect. Thus, while the Taiwanese genetic component disappeared when the Austronesian expansion ended at Easter Island, Austronesian languages survived. For linguists this is an important observation for understanding language variation. Language continuity is sometimes maintained despite population replacement.

3.5. The C1b-M347 Mutation and the Colonization of Australia.

The reader is invited to review Supplementary Figure 6.1. As detailed above, the founding populations of Sunda and Sahul had the C1-B477 mutation. Downstream variants of C1-B477, the C1-M38 "Green Cluster" and C1-M347 "Blue Cluster" mutations, later evolved among geographically isolated populations. As previously noted the C1-M38 mutation represents a founding lineage for populations that colonized Island Southeast Asia. C1-M347, on the other hand, represents a mutation that evolved among the aboriginal people of Australia. The most comprehensive study of indigenous Australian Y-chromosome genetic variation (Nagle et al. 2016a) reports an overall frequency of around forty percent within this population. Furthermore, the same study suggests that C1-M347 evolved from C1-B477 about 44 thousand years ago.

The discovery of the Australian-specific C1-M347 mutation was initially reported in 2007 by Hudjashov et al. In this report researchers utilized the enhanced resolution of downstream variation within the main haplogroup C-M130 marker to address a study (Redd et al. 2002) that reported Holocene geneflow between India and Australia about 10 thousand years ago. This study based their findings on a type of genetic marker called Short Tandem Repeats (STR's). Hudjashov et al. (2007) disagreed with the 2002 study and asserted that the Australian aborigines had not experienced any outside geneflow for a period of roughly 45 thousand years, from time that the continent was initially colonized by modern humans until the arrival of Europeans in the late eighteenth century (see, also, Nagle et al. 2016a).

A complete discussion of indigenous Y-chromosome variation among the Australian aborigines also requires a discussion of haplogroups M-P256, and S-B254. Accordingly this topic continues in Chapter 13: Section 6.

Section 4. Bi-directional Expansion of C1a-CTS11043 across Eurasia via a Northern Route.

As explained above, downstream variants of C1b-F1370 are important mutations for understanding the human colonization of India, Island Southeast Asia, and Australia via a southern route during Marine Isotope Stage 3. We now focus on C1a-CTS11043, the sister clade of C1b-F1370. At this point the reader may want to review Supplementary Figure 6.2. As shown by the figure, downstream from C1a-CTS11043 are the C1a-M8 and C1a-V20 mutations. Based on data from Poznik et al. (2016: Supplementary Table 10), both mutations evolved about 44 thousand years ago.

As mentioned previously in Chapter 4: Section 4, the D-M55 and C1a-M8 mutations stand as the genetic relics of the human colonization of Japan roughly 30 thousand years ago. Data from Sato et al. (2014) suggest that about six percent of contemporary Japanese have the

C1-M8 mutation. Furthermore, C1a-M8 is a Japanese-specific mutation, meaning that it is not found elsewhere, at least among contemporary populations (e.g. Hammer et al. 2006).

Ancient DNA results indicate a close phylogenetic relationship between modern Japanese with the C1a-M8 mutation and individuals who colonized Europe during Marine Isotope Stage 3. This is surprising because haplogroup C-M130 is rarely found among contemporary Europeans. We know that C1a-M8 and C1a-V20 were part of the genetic inventory of Paleolithic Europeans based on ancient DNA data acquired from individuals that died between 13 thousand and 35 thousand years ago. Part of the ancient DNA data stems from a 2016 report by Fu et al. This study reports a sample taken from remains found in Belgium. The sample belongs to C1a-CTS11043 and comes from the Goyet Q116-1 man who died about 35 thousand years ago. Another sample from the study, the Vestonice man, comes from the Czech Republic and belongs to C1a-V20. This individual died about 30 thousand years ago. The C1a-V20 mutation was also found in Russia. The Sunghir 1 man, who died about 34 thousand years ago near present-day Moscow (Sikora et al. 2017), has this mutation. Finally, C1a-P121, which is downstream from C1a-M8, was found in remains in Spain that date to around 13 thousand years ago (Villalba-Mouco et al. 2019).

Ancient DNA data from Paleolithic Europeans and East Asians should not be used to define a close genetic relationship between contemporary Europeans and Japanese. Rather the data supports a "northern migration" route during Marine Isotope Stage 3. In other words, about 50 thousand years ago the human tribe in the Levant split into two different groups. One group followed a southern route or a migration along the coastline of southern Asia. The genetic relics of this migration are the D-M55 and C1b-F1370 mutations. Another group migrated northwards from the Levant. Somewhere in Eastern Europe or Central Asia, another split occurred. Some traveled west in the direction of contemporary Belgium, and the other group traveled eastwards in the direction of contemporary Japan. The genetic relics of this bi-directional northern expansion include C1a-M8 and C1a-V20.

Section 5. The Importance of *C1* Mutations for Linguists.

The aboriginal Australians remained isolated from the rest of the world until about two hundred years ago. Thus the Australian language family has roots that extend to the out-of-Africa exodus. Taking this a step further, the C1b-M347 mutation, which is only found only among aboriginal Australians, supports the position that language evolved at least 100 thousand years ago. This follows the initial out-of-Africa migration and makes a huge (but plausible) assumption that the out-of-Africa tribe already had language. Turning now to C1-M38, this mutation represents an important genetic tool for deciphering the evolution of Papuan languages and the spread of Austronesian across the Pacific. The diversity of languages on New Guinea, as represented by the Papuan macro-family, is partly explained by the age of the C1b-M38 mutation. The C1b-M208, which is downstream from C1b-M38, resents an Island Southeast Asian component of Austronesian languages. Focusing now on C1b-M356, this mutation helps to explain linguistic diversity in India by defining indigenous and non-indigenous components of the gene pool. Finally, mutations downstream from C1a-CTS11043 support northern bi-directional migrations across Eurasia during Marine Isotope Stage 3. These migrations represent an important component of the *mammoth steppe hypothesis*, a discussion that surfaces in Chapters 14, 16, and 17. This hypothesis, in turn, helps to explain language variation in Eurasia and the Americas.

Section 6. Overview of the *C2* Mutations.

As previously detailed in Section 2, the C-M130 main haplogroup has two main branches, C1-F3393 and C2-M217, or alternatively, *C1* and *C2.* Both diverged from the C-M130 main haplogroup about 45 thousand years ago, during Marine Isotope Stage 3. Perhaps this occurred when the out-of-Africa human migration reached northern India. C1-F3393 then expanded rapidly across the Eurasian landmass. Those with C2-M217, on the other hand, appear to have "nested" in south central Siberia. Several thousand years later, after the Last Glacial Maximum, *C2* then expanded in the direction of East Asia, and then northwards into the Americas. The current distribution of C2-M217 and its variants now present a very useful tool for interdisciplinary analysis of the so-called *Transeurasian hypothesis* as presented in Section 7 (below). Additionally, the moderate frequency of C2-M217 found among Han Chinese, as discussed in Section 8, helps to decipher the evolution of Sino-Tibetan languages. Finally, *C2* mutations help to decipher the evolution of Native American languages (Section 9).

Section 7. Altaic and Transeurasian.

7.1 Overview.

Striking lexical and grammatical similarities found among the Japonic, Koreanic, Turkic, Tungusic, and Mongolic languages (e.g. Robbeets 2008) have been a topic of intense interest among linguists. The *Transeurasian hypothesis* has been formulated to explain these similarities (e.g. Robbeets 2017a). An approach to this hypothesis from the perspective of historical linguistics would classify these language families as part of an Altaic or Transeurasian macro-language family (or macro-phylum). As such, linguistic similarities are explained by the evolution of Japonic, Koreanic, Turkic, Tungusic, and Mongolic from a common proto-Altaic or proto-Transeurasian language. An alternative socio-linguistic approach to the Transeurasian hypothesis would view Japonic, Koreanic, Turkic, Tungusic, and Mongolic as part of a northeast Asian *Sprachbund*. As such, linguistic similarities stem from close geographical proximity and borrowing that has evolved over a prolonged period of time due to intense contact between the speakers of these languages.

At the Max Planck Institute for the Science of Human History, Dr. Martine Robbeets currently leads a project that explores the origins and expansion of the so-called Transeurasian languages: (http://www.shh.mpg.de/102128/eurasia3angle_group). The informal title for this the project is "Millet and Beans, Language and Genes," which reflects willingness among the researchers to employ multi-disciplinary perspectives in an effort to resolve a long-standing linguistic controversy. In a recent paper from 2017, Dr. Robbeets provides her views on the origins of Transeurasian from linguistic, archeological and genetic perspectives. One reason for citing this paper is that Dr. Robbeets suggests that haplogroup N-M231 is an informative for exploring the Transeurasian hypothesis. The evidence suggests, however, that N-M231is not a significant marker for Transeurasian languages, but rather for Uralic languages (see Chapter 14: Section 4.1). From a Y-chromosome perspective, C2-M217 and its downstream variants represent the markers of choice for exploring the *Transeurasian hypothesis*.

7.2. Origins of the C2-M217 Expansion.

The currently known C2-M217 internal phylogeny consists of four main lineages or clusters: the C2-P39 "purple cluster," the C2-M48 "red cluster," the C2-F1918 "green cluster," and the C2-M407 "blue cluster." C2-P39 is found among Native Americans. The remaining three represents genetic diversity in East Eurasia. These four clusters evolved roughly fourteen thousand years ago, at the onset of the Holocene. See Zhong et al (2010) for C2-M407; Wei et al. (2017b: Supplementary Figure S1) for C2-F1918; Karmin et al (2015) for C2-M48; Malyarchuk et al. (2011) for C2-P39. Additionally, available frequency data for C2-M48, C2-F1918 and C2-M407 among Turkic, Tungusic and Mongolic speaking populations reflect potential language contact among these groups. As this point the reader is directed to Supplementary Figure 6.4 and Supplementary Tables 6.2, 6.3 and 6.4 for additional information.

As explained in the previous paragraph, C2-M217 expanded during the Holocene. The next question seeks to identify the geographic origins of the expansion. Accordingly, a discussion of Ice Age refugia is necessary in order to provide important background information that helps to evaluate the Transeurasian hypothesis. The term "refugia" is the plural form of "refugium." For the purposes of this present discussion, both terms carry a discussion of where human populations congregated during the Last Glacial Maximum. As the reader may recall from Chapter 4: Section 1, roughly 50 thousand years ago, during Marine Isotope Stage 3, the weather across southern Asia improved. This facilitated a rapid expansion of the human tribe from the Levant to Asia and Australia. However, the weather deteriorated and the ice glaciers eventually reached their maximum southern extent in the northern hemisphere about 27 thousand years ago, a point that roughly equates to the fortieth northern parallel (see, e.g., Clark 2009). Geologists and earth scientists commonly refer to this event as the Last Glacial Maximum, a final and dramatic conclusion to a long Ice Age. The advance of ice glaciers curtailed human migration and populations expansions. Human populations settled in several refugia across the Eurasian landmass where they waited for better weather.

Once the glaciers retreated some populations, such those as in present-day Japan and Australia, remained in-place. Populations in other refugia expanded with the onset of the Holocene. For geneticists, the isolation of populations during the Last Glacial Maximum, and their subsequent post-glacial expansion during the Holocene, or alternatively, their continued isolation, represents a partial explanation for global genetic diversity. For linguists, this provides a partial explanation for global linguistic diversity. A study from 2016 by Gavashelishvili and Tarkhnishvili used computer simulation to identify the refugia where human survived during the Last Glacial Maximum. Additionally, they identified the human Y-chromosome haplogroups that expanded from these refugia with the onset of the Holocene. Their model was constructed utilizing a synthesis of climate, terrain, and hydrographic data, as well as data from fossilized pollen and plant remains. Data from Gavashelishvili and Tarkhnishvili (2016) place one of these refugia in the vicinity of south central Asia.

The paleo-climatological, anthropological, and genetic evidence suggest that C2-M217, along with variants of Q-M242 (see Chapter 16) and R-M207 (see Chapter 17), expanded from the same refugium in south central Siberia after the Last Glacial Maximum. This refugium, the so-called Altai-Sayan region, is located where China, Russia, Kazakhstan and Mongolia converge on a map. This area, along with much of Central Asia, has long been characterized by low precipitation and a vast stretch of prairie or "the steppes." These characteristics provide an explanation as to why this region may well have become an Ice Age refugium. Dolukhanov (2003) suggests that during the Last Ice Age, northern and central Europe were depopulated because of thick layer of ice and snow. However, the Central Eurasian steppes were covered by just a thin layer of snow because of low precipitation in the region. As such, the steppes provided an ideal habitat for a variety of large mammals including mammoths, woolly rhinoceros, wild horses, and bison. Even during the winter months these animals could easily forage as they simply had to scrape away a thin layer of snow to access the grass underneath. The Ice Age hunter-gatherers, in turn, hunted these large mammals that thrived in the region, and feasted on an abundant source of protein that could be harvested at a comparatively small expenditure of energy.

Y-chromosome data provided by Zhabagin et al (2017) may also support a south central Siberian refugium for haplogroups C2-M217, Q-M242 and R-M207. This study analyzed 780 samples from the nearby Central Asian region of Transoxiana. Source populations for the data include Kazakhs, Uzbeks, Turkmen, Dungan and Karakalpak. According to data provided by Zhabagin et al (2017), among the populations of the region, C2-M217 attains an overall frequency of thirty-one percent, R1a1a-M198 attains sixteen percent, and Q-M242 attains thirteen percent.

7.3. Transeurasian and Agriculture.

In her 2017 paper, Robbeets also suggests that the origins and expansion of Transeurasian languages follow the evolution and expansion of millet cultivation that began about eight thousand years ago at Xinglonggou in Inner Mongolia. Nevertheless, it would appear as though millet cultivation carries only part of the story that explains the evolution of some of the Transeurasian languages. Another important crop is rice. This explains the high population density observed in Korea and Japan. According to Stevens and Fuller (2017) millet and rice cultivation began in China roughly eight thousand years ago. From southeastern Manchuria the cultivation of foxtail and broomcorn millet eventually spread from China to Korea about 5.5 thousand years ago, followed by the spread of rice cultivation about 3.5 thousand years ago. Around 2.5 thousand years ago, millet and rice cultivation spread from Korea to Japan.

The story of agriculture in Central Asia is also an important factor in the evolution of Altaic languages (Turkic, Mongolic, and Tungusic). In this region, agriculture began about 5.5 thousand years ago when the horse was first domesticated, which appears to have occurred north-central Kazakhstan (e.g. Frachetti 2012). As mentioned earlier, horses were one of several large mammals hunted by prehistoric peoples who lived in the south central Siberian refugium around the time of Last Glacial Maximum. Thus the domestication of the horse should be seen as an effort to ensure its continued availability as a source of food. The adaptation of this animal as a means of transport occurred later when people began to ride horses. Soon thereafter, horses were used as draft animals to pull wagons and chariots.

It should also be emphasized that horse domestication represents only part of the success of agriculture in Central Asia. Around 4.5 thousand years ago cattle, goats and sheep appeared in the region and became part of the food economy. Then another important step in the evolution of Central Asian agriculture occurred shortly thereafter, about four thousand years ago, when mobile pastoralists began to cultivate crops such as millet, barley and wheat. China was the source of millet that was initially grown in Central Asia (Stevens and Fuller 2017). Barley and wheat, as well as goats and sheep, however, came from the Middle East (Bellwood 2005: 84-86; Spengler et al. 2014).

Robbeets (2017a) suggests that an expansion of millet cultivation from China brought Transeurasian languages to Central Asia. It is difficult to find genetic support because the internal phylogeny of C2-M217 requires greater clarification. However, according to the archaeological record by around three thousand years ago nomadic pastoralism spread from Central Asia to Mongolia (Askarov et al. 1992). Perhaps this expansion carried Transeurasian languages to East Asia.

7.4. C2-M217 and Turkic.

As noted earlier, the *Transeurasian hypothesis* seeks to explain the origins of Transeurasian languages. One of the languages families within this macro-family classification is Turkic. The reader is now invited to examine Supplementary Table 6.5, which presents a survey of Turkic-speaking populations that have appeared in published Y-chromosome studies in the last 20 years. The table illustrates the fact that Turkic-speaking populations appear over a wide geographical expanse. Examples include the following: Turks in Southwest Asia; Azerbaijani in the Caucasus; Kazakhs, Kyrgyz, Turkmen, and Uzbeks in Central Asia; Ainu (Änyu) of East Asia; and Yakuts of Siberia. Where and when Turkic languages evolved appears to still be very much a mystery (e.g. Kornfilt 2009). However, Orkun Inscriptions found in Mongolia and Old Uyghur manuscripts found in Xinjiang, China from about the eighth or ninth century, point to East Asia.

At this point the reader is directed to Supplementary Table 6.6 which reports the frequency of C2-M217 among several different Turkic-speaking populations. Among the Central Asians Kazakhs, the frequency of C-M217 is very high. Based on the high frequency of C2-M217 among the Kazakhs, as well as the distribution and frequency of C2-M217 throughout East Eurasia, the genetic evidence potentially identifies Central Asia as a potential homeland of Turkic languages. However, the internal phylogeny of C2-M217 still remains a mystery because almost the effort devoted to refining downstream C2-M217 markers involves those that likely evolved in Mongolia or Northeastern China: the C2-M48 "red cluster," the C2-F1918 "green cluster," and the C2-M407 "blue cluster" mutations (see, also, Supplementary Figure 6.4 and Supplementary Tables 6.2, 6.3 and 6.4).

Zerjal et al. published a study in 2003 that claimed to have found a unique Y-chromosomal haplogroup C-M130 lineage based on unique pattern of Short Tandem Repeats (STRs). The study reports that this lineage was found among sixteen of Central Asian populations with an overall frequency of around eight percent. According to the study, the lineage evolved in Mongolia about 1,000 years ago and was spread to Central Asia by Genghis Khan and his descendants. This lineage became known in the literature as the "Genghis Khan Y-profile" or "Genghis Khan star-cluster" (e.g. Abilev et al. 2012). However, Wei et al. (2017b) determined that the C2-F1918 haplogroup is the mutation that was previously identified in the literature as the Genghis Khan star-cluster. The researchers further report that C2-F1918 evolved about fifteen thousand years ago, a date that clearly predates the Mongol Empire, which, in turn, undermines the purported reproductive success of Genghis Khan.

Abilev et al. in their 2012 study, based on their analysis of the "Genghis Khan star-cluster," report that around seventy-six percent of the Kazakh Kerey tribe have this mutation. The Kerey are the largest of the Kazakh tribes. According to the study, the Keraits, a Mongolic tribe, were defeated by Genghis Khan. Many escaped and joined the Turkic people. They adopted Turkic language and the term "Kerey" is a Turkic form of "Kerait." The study further reports that many of the contemporary Turkic ethnic groups evolved from Keraits, including Tatars, Karachays, Nogays, Bashkirs, Kazakhs, Uzbeks, Kyrgyz, and Altaians.

About thirty percent of all Turkic-speakers in the world reside in modern-day Turkey (e.g. Kornfilt 2009). The largest Y-chromosome survey of Turkey (Cinnioglu et al 2004) indicates that less than one percent of Turkish males have the haplogroup C-M130 mutation. This agrees with the historical record and follows the demise of the Byzantine Empire. Thus language shift occurred in Anatolia without significant admixture with Turkic-speakers from Central Asia or Northern Eurasia. This underscores the following: language expansion can occur in the absence of a population expansion. Thus language shift appears to partially explain the expansion of Turkic language. The Yakuts, a Turkic-speaking population of Siberia, provide yet another example. Their reliance on reindeer herding and the high frequency of N-M231 suggest that they initially spoke a Uralic language (e.g. Pakendorf et al. 2006).

7.5. C2-M217 and Mongolic.

Another Transeurasian language is Mongolic. *Ethnologue* (2017) classifies thirteen languages within the Mongolic language family. Twelve of the languages are spoken either in China, Russia or Mongolia. The other Mongolic language, Mogholi, is found in Afghanistan. Arguably, the earliest attestation of Mongolic languages are the so-called “Para-Mongolic” Khitan scripts dating to about the tenth century (Kane 1989: 11-37; Janhunen 2003a: 394-396), which prepared under the auspices of the Liao Dynasty. Pre-Classical Mongolic texts later emerged during the reign of Genghis Khan in the thirteenth century (Janhunen 2003b: 32-33).

At this point the reader is directed to Supplementary Table 6.7 which provides a survey of Mongolic-speaking populations living in Russia, Mongolia and China. As shown by the table, C2-M217 attains a very high frequency among some Mongolic-speaking populations, similar to what is observed among speakers of Turkic languages (see Section 7.4) and Tungusic languages (see Section 7.6). The reader is also directed to Supplementary Tables 6.2, 6.3 and 6.4 which report C2-M407, C2-M48 and C2-F1918 variation among Turkic, Mongolic and Tungusic-speakers. In a recent study from 2017, Huang et al further refined the phylogeny of C2-M407 (see, also, Supplementary Figures 6.4 and 6.5). According to the study, C2-F8465, a downstream variant of C2-M407, represents the genetic signature of Mongolic languages. The study further reports that this mutation evolved roughly four thousand years ago in Northeast Asia. Thus, unlike Turkic, the putative homeland of Mongolic languages seems much clearer.

7.6. C2-M217 and Tungusic.

Tungusic is another Transeurasian language. According to *Ethnologue* (2017), the Tungusic language family consists of eleven languages spoken by around 55 thousand speakers either in Northeastern China or Eastern Siberia. Tungusic languages include those spoken by the Even and Evenki people. These closely related ethnic groups consist of small populations in Siberia whose survival strategy once included the domestication of reindeer. In contrast, another Tungusic language, Manchu, stands as a former linguistic heavyweight, a relic of the Qing Dynasty of China. However, the Qing Dynasty eventually collapsed in 1912, and as a result, the Manchu language rapidly became moribund.

The reader is invited to examine Supplementary Table 6.8 which provides a survey of Tungusic-speaking populations and reported frequencies of C2-M217. Like Turkic and Mongolic, these populations also exhibit a high frequency of the mutation. The reader is also invited to examine Supplementary Tables 6.2, 6.3 and 6.4 which provide frequency data for the known C2-M217 Holocene expansion markers. Based on these data, language contact between proto-Turkic, proto-Mongolic and proto-Tungusic populations seems possible. Nevertheless, small population size and the associated phenomenon of genetic drift limit the ability of genetic markers as a tool for identifying the geographic origins of Tungusic languages (e.g. Duggan et al. 2013). The earliest attestation of Tungusic stems from texts that appeared in the twelfth century. Under the auspices of the Jin Dynasty, these texts were written in the Jurchen language using characters borrowed from Khitan (a Mongolic language) and Chinese (Kane 1989:1-10). These texts, along with heavy frequencies of N-M231 found in the Tungusic populations of Siberia (e.g. Pakendorf et al. 2007; Fedorova et al. 2013) point to northeastern Asia as the putative homeland of Tungusic.

Variants of haplogroup N-M231 represent the genetic signature of Uralic speakers and reindeer herders (see Chapter 14). Thus it would appear that Tungusic speakers with C2-M217 migrated northwards from East Asia into Siberia sometime in the prehistoric past. Overtime, populations like the Even and Evenki began to herd reindeer, a subsistence strategy that is well suited to the Siberian climate. Admixture with Uralic-speaking population, as well as genetic drift, eventually produced the high frequency of N-M231 found in some Tungusic-speaking populations in Siberia (see Karafet et al. 2002 for a detailed discussion).

7.7. C2-M217 and Koreanic.

Koreanic represents another Transeurasian language. The Koreans stand among the ancient ethnic cultures of the world. A good starting point for discussing their ethnogenesis may well be the beginning of the Jeulmun pottery period about ten thousand years ago. However, Kim (2009) suggests that a reliable attestation of the Korean language emerged comparatively late in the Korean history, about six hundred years ago, when the Korean hangul script was introduced in a document called the *Hunminjeongeum*. According to the same source, classification of the Korean language has been difficult. The so-called "southern theory" attempted to associate Korean with Dravidian or Austronesian. The northern theory, on the other hand, classified Korean as part of an Altaic macro-family.

Contemporary linguistic classification of Korean has generally disassociated the language with Altaic. In their 2014 seventeenth edition, *Ethnologue* classified Korean as a language isolate. However, with the eighteenth edition, which was released in 2015, *Ethnologue* re-classified Korean within a newly created language family called Koreanic. This language family contains just two languages, with Korean having, by far, the largest number of speakers, which totals 48 million on the Korean peninsula, and 77 million worldwide. Jejueo, the other Koreanic language, has just five thousand speakers on Jeju Island in the Korean Straights.

At this point the reader is directed to Supplementary Table 6.9. As shown by the table, around fifteen percent of Koreans have the C2-M217 mutation. The best refinement of C2-M217 variation among the Koreans emerged in 2015 with the study published by Kwon et al. At this point the reader is directed to Supplementary Figure 6.5 and in particular, the mutations surrounded by a green border, which were reported in the study that was just cited. As shown by the figure, C2c-F1067 seems to unite the genetic history of Koreans with that of central Eurasia. Of course the genetic history of Koreans is not complete without a discussion of haplogroup O-M175. The reader is directed to Chapter 15: Section 14, for additional information.

7.8. C2-M217 and Japonic.

The Transeurasian macro-family includes Japonic. Like Korean, the Japanese language has also proven difficult to classify. In the past, some linguists have placed this language within the Altaic super-family, along with Korean, Mongolic, Turkic and Tungusic (e.g. Shibatani 2009). *Ethnologue* (2017) currently places Japanese within the Japonic language family. The Japonic family has two main branches, the Japanese language, which is spoken by over 127 million people throughout Japan, and the Ryukyuan branch, which contains eleven languages, is spoken on the island of Okinawa. As previously detailed in Chapter 4: Section 4 and Section 4 of this present chapter, the starting point for a discussion of Japonic languages begins roughly 30 thousand years ago with the initial human colonization of the present-day Japanese islands. As explained in these sections, the prehistoric Jomon people represent the cultural relic of this Paleolithic migration. Haplogroups D1b-M55 and C1-M8, on the other hand, provide the genetic artifacts.

The Neolithic Yayoi culture represents the second and last major human migration that settled in present-day Japan. Downstream variants of the O-M175 main haplogroup carry the bulk of the genetic evidence for this event. This genetic evidence supports the anthropological perspective that places the origins of Yayoi culture in Korea (see Chapter 15: Section 15 for additional details). Addition genetic evidence comes from C2-M217 mutations that are found in about seven percent of Japanese males (see Supplementary Table 6.10). A comparison of downstream variants within C2-M217, as reported by Naitoh et al. 2013 and Kwon et al. 2015 (see, also, Supplementary Figure 6.5), point to Korea as the source of C2-M217 variation in Japan. Thus C2-M217, like O-M175, stands as a genetic relic of the Yayoi culture.

7.9. Analysis of the Transeurasian Hypothesis.

Millet cultivation represents only a small part of the story of Transeurasian languages. The ancestral populations of contemporary groups that now speak Transeurasian languages survived and thrived due to complex combination of factors that may have begun with successful adaptation to climate change during the Last Glacial Maximum. Turning now to the Altaic component, these languages thrived and survived because of mobile pastoralism, the successful domestication of the horse in Central Asia, the successful domestication of reindeer in Northern Eurasia, the adoption of sedentary agriculture in some regions, and the expansion and demise of nomadic societies such as the Mongol Empire. Of course, the phenomenon of language shift stands as another factor that should not be neglected. Turning now to the evolution of Koreanic and Japonic, the genetic data, along with the archaeological and historical record, reflect that unlike the Altaic languages, geographical and cultural isolation played a substantial role in the evolution of both language families. Furthermore, rice cultivation clearly distinguishes the evolution of Koreanic and Japonic with that of Altaic. Korean and Japanese now occupy a huge corner of the global linguistic tapestry because their ancestors found a survival strategy that supports a very high population density.

Section 8. Han Chinese and C2-M217.

Ethnologue (2017) classifies Chinese as both a macro-language and as a branch within the Sino-Tibetan language family. With over 1.2 billion speakers, it goes without saying that

Chinese plays a significant role within the global tapestry of linguistic diversity. The best source of genetic data for exploring the evolution of the Chinese macro-language comes from the Han Chinese. They are, by far, the largest ethnic group in China, representing almost ninety-two percent of the population (CIA World Factbook).

At this point the reader is directed to Supplementary Table 6.11 which provides C2-M217 frequency data for the Han. Based on this table and data extrapolated from Zhong et al. (2011), C2-M217 attains an overall frequency of about twelve percent among this population. However, based on a review of the published data, downstream C2-M217 variation among the Chinese still remains very much a mystery. We know that the Central Asian contribution is very minimal. Based on data from Wei et al (2017b: Table S1), the frequency of Central Asian C2-M217 mutations (C2-M48, C2-M407 and C2-F1918) among the Han is virtually non-existent. Thus, researchers should look elsewhere for the genetic mutations that carry the story of Sino-Tibetan language. At this point the reader is directed to the discussion in Chapter 15: Section 3.

Section 9. Native Americans and C2-M217.

C2-M217 and haplogroup Q-M242 (see Paper 5.16) mutations represent important genetic tools for deciphering the prehistory of Native American languages. Among the indigenous populations of North America, haplogroup Q-M242 carries about ninety-three percent of the indigenous genetic component, whereas C2-M217 represents the remaining seven percent (e.g. Zegura et al. 2004). However in South America Q-M242 represents almost all of the indigenous Native American genes (Geppert et al 2011; Roewer et al. 2013; Jota et al. 2016). Here, C2-M217 is extremely rare among the indigenous populations. Pinotti et al. (2019), for example, suggest that C2-M217 has only been found in thirteen indigenous South Americans. The same study also reports that indigenous South American have a C2-M217 variant that is evolutionary distant from the C2-M217 variant found among the indigenous peoples of North America, the C2b-P39 mutation. According to Pinotti et al. (2019) the unique South American C2-M217 variant and the unique North American C2-M217 variant (C2b-P39) diverged from a common ancestor roughly 22 thousand years ago.

The reader is now invited to examine Supplementary Figure 6.4. The C2-P39 mutation is highlighted by a purple border. As shown by the figure, C2-FGC28881.2 is a phylogenetic sister clade of C2-P39. This mutation was reported by Wei et al. in 2017b. According to the study, C2-FGC28881.2 forms part of the gene pool of contemporary Koryaks. Among the Paleo-Siberian peoples of Asia, Koryaks have traditionally lived along the Bering Sea near the Kamchatka Peninsula. They speak a language belonging to the Chukotko-Kamchatkan language family. Moreover, they have traditionally employed a hunter-gather subsistence strategy that included the harvesting of whales.

The above discussion of Koryaks serves a linguistic purpose which involves the bi-directional movement genes and culture across the Bering Sea. Such a discussion requires additional cultural context that follows the evolutionary history of the Q-M242 haplogroup among the indigenous peoples of Alaska. Accordingly, a discussion of Koryaks and C2-FGC28881.2 continues in Chapter 16: Section 8.

Section 10. The Importance of *C2* Mutations for Linguists.

Based on frequency data, the C2-M217 mutation represents a very important tool for understanding the evolution of the so-called “Altaic” macro-family of languages, which consists of the Turkic, Tungusic, and Mongolic language families. Additionally, C2-M217 attains a moderate to low frequency among the Japanese and Koreans. Thus the mutation helps in the analysis of the so-called Transeurasian hypothesis which advocates a common origin for the Japonic, Koreanic, Turkic, Mongolic and Tungusic language families. C2-M217 also attains a moderate frequency among the Han. However, unlike the Transeurasian languages, Central Asian variants of C2-M217 are almost non-existent among the Han. This suggests that C2-M217 is not an especially informative marker for the Sino-Tibetan language family. Finally, C2-M217 represents a useful marker for deciphering the origins of linguistic diversity among the native people of both Siberia and the Americas.

Section 11. Summary of C-M130 Data.

The evolutionary history of the *C1* and *C2* branches is so vastly different that one wonders if C1-F3393 and C2-M217 are, in fact, main haplogroups. C1-F3393 represents population expansions during the Paleolithic whereas C2-M217 represents population expansions during the Holocene. Perhaps C-M130 represents a higher level evolutionary step, or paragroup, within the Y-chromosome phylogeny.

Chapter 7: Haplogroup G-M201.

Section 1. Overview of G-M201.

At this point the reader is directed to Supplementary Figure 1.1 from the first chapter. DR-M168 represents the ancestral mutation of Y-chromosome haplogroups that evolved outside of Africa. As previously detailed in Chapter 4 to 6, Haplogroups D-M174, E-M96 and C-M130 evolved in the Middle East. The sister clade of C-M130, the FR-M89 mutation, eventually evolved into G-M201 and HR-M578 around forty-six thousand years ago (Poznik et al. 2016: Supplementary Table 10). G-M201, like D-M174, E-M96 and C-M130, also evolved in the Middle East (e.g. Rootsi et al. 2012). D-M174, E-M96 and C-M130 expanded out of Southwest Asia during the Pleistocene, about fifty thousand years go. G-M201, on the other hand, began to expand out of this region much later, roughly ten to twelve thousand years ago, a period that coincides with the evolution of agriculture in Southwest Asia (See Rootsi et al 2012: Main Report and Supplementary Table 4).

As shown by Supplementary Table 7.1, the frequency of the G-M201 peaks in the Caucasus and then tapers off to less than ten percent elsewhere. It should be noted that although G-M201 attains a heavy frequency among some of the Kazakh tribes, overall G-M201 frequencies in Central Asia are, nevertheless, low (i.e. Zhabagin et al. 2017). Turning now to the internal phylogeny of G-M201, within this mutation one finds two main branches, G2-P287 (see Supplementary Figure 7.1) and G1-M285 (Supplementary Figure 7.2). Turning now to the data tables, as shown by Supplementary Table 7.2, the distribution of G1-M285 is rather limited and is confined almost exclusively to populations in Asia. Supplementary Table 7.3, on the other hand, indicates that the distribution of G2-P287 is much broader, having a range that extends from Western Europe to Central Asia.

As noted previously, G-M201 and its variants expanded out of Southwest Asia during the Neolithic. The evolution of agriculture in Southwest Asia was previously introduced in Chapter 5: Section 2. As detailed in this discussion, the literature often pinpoints the so-called "Fertile Crescent" as the homeland of agriculture within this region. About fourteen thousand years ago people initially harvested wild cereals. This led to a series of innovations that included the development of pottery, the genetic modification of cereals and legumes for cultivation, and the domestication of goats and sheep. As detailed in Chapter 5, agriculture in Southwest Asia eventually spread into North and East Africa roughly 6.4 thousand years ago. Furthermore, as mentioned in Chapter 6: Section 7.3, about 4.5 thousand years ago elements of Southwest Asian agricultural package also spread to the Central Asian steppes.

Turning now to Europe, South Asia, Central Asia, and the Caucasus, the Neolithic transformation in all four regions resulted from an expansion of agriculture from Southwest Asia. Within these regions, downstream variants of G-M201 now stand as the genetic relic of this transformation. For linguists this observation facilitates three different discussions. First, G-M201 variation supports the *early farming dispersal hypothesis* (Bellwood 2005:1-11). According to the hypothesis the current distribution of many language families throughout the world follows the initial expansion of early agriculture, an innovation that evolved independently in several regions of the world. Taking this a step further, within the context of Indo-European languages the initial expansion of this language family from the Middle East to Western Europe and India follows the westward and eastward Neolithic expansion of agriculture from Southwest Asia that began about nine thousand years ago (Bellwood 2005: 67-97; 201-207). Surprisingly, within a South Asian context, the *early farming dispersal hypothesis* may also explain the expansion of Dravidian languages from Pakistan to southern

India. This is based on the observed frequency of G-M201 as found in the Brahui of Pakistan and that found among the Dravidian-speaking farmers of Southern India.

The second discussion raised by the distribution of G-M201 variation is that these data undermine a longstanding hypothesis that associates the spread of Indo-European with an expansion of steppe nomads from Eastern Europe or Central Asia (e.g. Gimbutas 1997; Anthony 2007; Anthony and Ringe 2015). More specifically, G-M201 demonstrates the absence of a major demographic Bronze Age migration from the steppes as advocated by proponents of this hypothesis. Of course, a major demographic expansion is not a prerequisite for language expansion as was the case with Turkish after the fall of the Byzantine Empire. However, such a scenario, especially for Indo-European, would imply that people over a vast expanse, from Western Europe to India, switched languages as the result of language contact with steppe nomads. In fact, Anthony (2008) proposes the status and prestige of steppe nomad culture mediated this switch in languages.

The *early farming dispersal hypothesis* (Bellwood 2005), on the other hand, provides a much more empirical explanation of how Indo-European spread over a vast geographical expanse. Europe and South Asia underwent the exact same cultural transformation at almost exactly the same time: the adoption of the Southwest Asian agricultural package beginning eight to nine thousand years ago. Of course another persuasive aspect of the *early farming dispersal hypothesis* is that it follows a trend observed throughout the world. Indo-European is just one of several different language families that co-expanded with the spread of early agriculture. The other language families include Arawak, Niger-Congo, Afro-Asiatic, Dravidian, Sino-Tibetan, Trans-New Guinea, Uralic, Austro-Asiatic and Austronesian. Thus, proponents of the *Kurgan hypothesis* must ask themselves the following question: Why should Indo-European be an exception?

Finally, the G-M201 mutation facilitates an examination of the complex tapestry of language variation in the Caucasus region. Within this compact region four different languages are represented: Indo-European, North Caucasian, Kartvelian, and Turkic. Haplogroup G-M201 helps to provide an explanation because the mutation attains an astonishingly high frequency among several populations in the region. As shown by Supplementary Table 7.1, among Indo-European-speaking Ossetians the frequency is about seventy percent. Among Georgians (Kartvelian languages) the frequency is around fifty percent. Among the Abkhaz (North Caucasian) the mutation attains a similar frequency. Finally, among Turkic-speaking populations, such as Balkars and Karachays, the frequency is also significant, close to thirty percent.

Section 2. The Neolithic in Europe.

Agriculture and Indo-European languages may well have co-expanded across Europe during the Neolithic. According to Bellwood (2005: 67-84), the expansion of agriculture from Southwest Asia to Europe follows three different trajectories. The first trajectory involves the maritime colonization of the Mediterranean islands. About ten thousand years, farmers from Anatolia (modern-day Turkey) settled on the island of Cyprus. From this location farmers later migrated to Crete, Corsica and Sardinia. From a Y-chromosome perspective, the G2-L91 mutation is a particularly strong genetic relic of these maritime expansions (See Supplementary Table 7.8; Rootsi et al. 2012; Francalacci et al. 2015; Voskarides et al. 2016).

The second trajectory involves the expansion of farming along the southern Mediterranean coast of mainland Europe, from Western Turkey to Portugal. This began about 8.5 thousand years ago and required about a thousand years. The third and final trajectory involves an expansion of farming that also began from Western Turkey. Here, farming expanded northwards through the Balkans and then westwards across Central Europe. This expansion also required about a thousand years and is often identified in the literature as an expansion of the Linear Pottery Culture (LBK). A particularly strong genetic relic of the LBK expansion is the G2-L497 mutation (see Supplementary Table 7.4). Other genetic relics include the G2-M406 mutation (see Supplementary Table 7.7) and G2-M527 (see Supplementary Table 7.6). See, also, Rootsi et al. (2012) and Berger et al. (2013).

Three additional points concerning the European Neolithic are worth mentioning. First, Europe received the full Neolithic package from Southwest Asia, which included pottery, cereals such as einkorn and emmer wheat, legumes such as lentils, and farm animals, such as sheep, goats, and pigs. Second, by 5400 BC the Linear Pottery Culture expansion had terminated at the coastal plain of Northern Germany and the English Channel in the Low Countries. The Neolithic transition in the British Isles required about another thousand years, and Scandinavia required an even longer period of time. Finally, variants of haplogroup G-M201 have been found in ancient DNA samples taken from human remains found in Europe at archeological sited dated to the Neolithic period (e.g. Haak et al. 2010; Lacan et al. 2011; Szecsenyi-Nagy et al. 2015; Fregel et al. 2017).

Section 3. The Neolithic in South Asia.

According to the *early farming dispersal hypothesis* (Bellwood 2005), the arrival of Indo-European Languages in South Asia follows the spread of agriculture from Southwest Asia during the Neolithic. It should be noted that the Neolithic in South Asia (contemporary Pakistan and India) saw the adoption of Southwest Asian crops such as wheat, barley, lentils, chickpeas, flax and linseed (e.g. Fuller 2006: 20). However, linguists should also note that besides a Southwest Asian component, the Neolithic transition in India also saw the adoption of crops from Africa, such as sorghum and cowpeas (e.g. Crowther et al. 2017), as well as rice from East Asia (see Chapter 15: Sections 2 and 11).

Focusing now on the Southwest Asian component of the South Asian Neolithic, the most likely path would have probably traversed the southern shore of Caspian Sea rather than traversing through the Iranian deserts. What is more certain is that by around nine thousand years ago numerous farming settlements appeared in Mehrgarh, which is found in the Balochistan region of Pakistan. Shortly thereafter, farmers penetrated the Indus Valley of Pakistan and western India. The Neolithic transition in the Indus Valley is often attributed to the Harappan culture in the literature. Over the course of several thousand years, elements of the Southwest Asian agricultural package eventually migrated eastwards from the Indus Valley into the Ganges Valley and southwards into southern India and Sri Lanka.

Haplogroup G-M201 records a weak but important signal of the Southwest Asian agricultural expansion among Pakistani and Indian populations. For Pakistan as a whole Sengupta et al (2006) report a frequency of around 4.6 percent. However, G-M201 frequencies appear somewhat stronger among Indo-Iranian speaking populations in Pakistan, such as the Kalash and Pashtuns (Di Cristofaro et al. 2013, Lee et al 2014). Turning now to India, Sengupta et al. (2006) report that G-M201 represents less than one percent of the population. Interestingly, the figure stands at around five percent in the Tamil Nadu region at the southern tip of India. This was reported in a study (Arunkumar et al. 2012) presenting data for over sixteen hundred men, many of whom are Dravidian-speaking farmers.

The unexpected frequency of G-M201 among the Dravidian farmers of Tamil Nadu, as reported by Arunkumar (2012) certainly reflects the potential of this haplogroup as an informative mutation for deciphering the evolution of Dravidian languages. The Dravidian language consists of eighty-five languages spoken by around 228 million people (*Ethnologue* 2017). Examples include Tamil, Malayalam, Kannada, and Telugu. It should be noted that Dravidian-speakers are mostly found in southern India. However, the Brahui people of Pakistan represent a distant linguistic island of two million Dravidian speakers in a sea of Indo-European-speakers. Like the Dravidian populations of southern India, haplogroup G-M201 also surfaces among the Brahui at a frequency of sixteen percent (see Di Cristofaro et al. 2013 and Supplementary Table 7.11). This raises an interesting question, whether the Indus Valley defines the putative source of Dravidian. The status of haplogroup G-M201 as a genetic marker of the South Asian Neolithic, the location of Brahui within the territory once held by Harappan culture, and the spread of the agriculture from this area into southern India, as told by the archaeological record, seem to support this position. Perhaps the Brahui are descendants of hunter-gatherers that admixed with Southwest Asian farmers during the Neolithic and adopted agriculture as a subsistence strategy while retaining Dravidian. The admixed population then migrated out of the Indus Valley along with Indo-European speaking populations.

It should be noted that variants of the J2-M172 mutation provide a stronger signal for the South Asian Neolithic. Accordingly, the discussion of language variation in South Asia continues in Chapter 10.

Section 4. The Neolithic in Central Asia.

The Central Asian Neolithic was previously introduced in Chapter 6: Section 9. On the Central Asian steppes, which stretch from Eastern Europe to Mongolia, mobile pastoralism has been the survival strategy for thousands of years. Agriculture in this region began about 5.5 thousand years ago in north-central Kazakhstan with the domestication of the horse. About a thousand years later, cattle, goats and sheep appeared on the steppes and became part of the pastoral food economy. Finally, about four thousand years ago, mobile pastoralists began to cultivate crops such as millet, barley and wheat. China was the source of millet. Barley, wheat, goats and sheep, came from Southwest Asia.

The evolution of mobile pastoralism on the Central Asian steppes represents a significant development for linguists as many believe that steppe nomads were responsible for initial dispersal of Indo-European languages (e.g. Anthony and Ringe 2015). Their position advocates a southward expansion of steppe nomads as the source of Indo-Iranian languages as found in Iran and South Asia. However, the archaeological record fails to support a mass southward migration of steppe nomads. Rather, mobile pastoralism in Central Asia represents a one-way expansion of farming from Southwest Asia. Additional archeological support for this position comes from Jeitun, an archaeological site in Turkmenistan near the Iranian border. This site represents the earliest expansion of agriculture from Southwest Asia into Central Asia, dating to about eight thousand years ago (Bellwood 2005: 84-86). Soon thereafter, agriculture appeared among the Hissar culture of Tadjikistan, at Kel’teminar in Kazakhstan, within the Ferghana Valley of Uzbekistan, and at Oshkona in Tadjikistan (Fuller 2006). Being that Jeitun lies a considerable distance south of where steppe pastoralism arose, researchers have evidence of a slow northward advance of the Southwest Asian agricultural package onto the Central Eurasian steppes.

In the Central Asian country of Afghanistan, the frequency of G-M201 among Indo-Iranian speaking Pashtuns and Tajiks suggests that this marker might record the history of Indo-European population expansions in this region (Lacau et al. 2012; Di Cristofaro et al. 2013; Lee et al. 2014). Furthermore, among the Pashtuns the G-M377 mutation stands as a particularly strong G-M201 variant (see Supplementary Table 7.9). Thus, Indo-Iranian populations in Central Asia may have descended from the Neolithic farmers that settled in Jeitun and the Indus Valley. Additional support for this position comes from Balanovsky et al. (2015). This study focused on the distribution of G1-M287 variation in Asia. Based on their analysis of the data they found that the expansion of Indo-Iranian languages correlates well with an expansion of agriculture from Southwest Asia, rather than a southward migration of steppe nomads. Furthermore, in Iran, which represents a transit point for the expansion of Indo-Iranian populations into Central and South Asia, G-M201 attains a frequency of almost twelve percent (Grugni et al 2012). Finally, within the Central Asian region of Transoxiana the G-M201 mutation has an overall frequency of about three percent and stands as the genetic relic of farmers that entered the region from the south during the Neolithic (Zhabagin et al. 2017).

Section 5. The Neolithic in the Caucasus.

The G2-P16 and G2-U1 mutations represent most of the G-M201 variants in the Caucasus (Rootsi et al. 2012). See, also, Supplementary Tables 7.5 and 7.10. The Caucasus region lies between the Black and Caspian Seas, and includes parts of Russia as well as Armenia, Azerbaijan, and Georgia. The Southwest Asian agricultural package arrived in the Caucasus region about 8.5 thousand years ago (Bellwood 2005: 85). Geneticists suggest that the arrival of agriculture in the region also brought populations with the haplogroup G-M201 mutation (e.g. Herrera et al. 2012; Rootsi et al 2012; Yunusbayev et al. 2012; Hovhannisyan et al. 2014; Karafet et al. 2016). However, the Caucasus region may well have been an Ice Age refugium for populations with G-M201 mutations (Gavashelishvili and Tarkhnishvili 2016), and as such, the original source population for the mutation potentially came from this region.

One interesting question that has surfaced from genetic studies of the Caucasus involves Indo-European languages. A recent study (Balanovsky et al. 2017b) extended the Caucasus region out onto the Armenian plateau of eastern Turkey, the homeland of the Armenian people. The study asserts that genetic variation in Southwest Asia follows a lowland/upland contrast. In the lowlands, which the study defines as the Levant, Mesopotamia and the Arabian Peninsula, one finds Afro-Asiatic. In the uplands, which the researchers define as the Anatolian, Armenian and Iranian plateaus, one finds Turkic and Indo-European. Turkic, of course, later migrated into the region. Indo-European, on the other hand, potentially originated in the Southwest Asian highlands. Given the close proximity of Armenians to this purported Indo-European homeland and the elevated frequency of G-M201 within this population, Armenians and haplogroup G-M201 seem to be part of the equation that defines the origins and spread of the Indo-European language family. On the other hand, in the Caucasus one also finds populations that speak North Caucasian and Kartvelian languages. Perhaps these languages are indigenous, and Indo-European represents languages that were imported into the Caucasus from Anatolia or the Levant. Furthermore, among North Caucasian and Kartvelian populations, haplogroup G also attains a substantial frequency similar to what is observed among Indo-European-speaking populations.

In an interesting paper from 2008 the linguist Bernard Comrie suggested that extreme linguistic variation found in the Caucasus reflects populations that have remained isolated because of topography and strict adherence to endogamy (marriage within the same group). From a genetics perspective, this isolation would suggest that genetic drift in the region has limited genetic variation, and as such has produced the high frequency of G-M201 among many of the populations of the region. Following now the cultural and geographic isolation as seen in the Caucasus, it would appear that such isolation defines the position of several language families within the global tapestry of languages, families that include North Caucasian and Kartvelian.

Section 6. Conclusions for G-M201.

Haplogroup G-M201 and its variants facilitate a discussion of the spread of Indo-European languages and the theoretical approaches to this problem. Another important marker for this discussion is J2-M172. Accordingly, this discussion continues in Chapter 10. However, a preliminary remark is in order here: wherever you find J2-M172, G-M201 is close behind. The distribution of G-M201 variation is, indeed, rather interesting and at times puzzling. For example, the G2-L30 variant is found in Judeo Tats, Bagvalal, and Nogais of the Caucasus region (Karafet et al. 2016). However, the same mutation is also found in Flanders (Larmuseau et al. 2014). Given the distances involved, G-M201 must have expanded very rapidly during the Neolithic, and this expansion suddenly stopped.

Besides Indo-European, G-M201 variation helps to decipher the prehistory of the Dravidian, Kartvelian, and North Caucasian language families. Potentially, the putative homeland of Dravidian languages is the Indus Valley. Kartvelian and North Caucasian represent indigenous languages of the Caucasus that never expanded out of this region due to cultural and geographic isolation.

Chapter 8: Haplogroup H-M2713.

Section 1. Overview.

In order to understand the phylogenetic history of the H-M2713 main haplogroup the reader should examine Supplementary Figure 1.1 from the first chapter. As shown by the figure, HR-M578 and G-M201 are sister clades. H-M2713 evolved from HR-M578. According to Poznik et al. (2016: Supplementary Table 10), this occurred around forty-six thousand years ago. The reader is now directed to Supplementary Figure 8.1 which outlines the internal phylogeny of H-M2713 and its informative variants. The internal structure contains two main divisions, H1-M3061 and H2-P96. According to Poznik (2016), both mutations evolved around thirty-seven thousand years ago.

In order to identify where the H-M2713 main haplogroup evolved, several points need to be discussed. First, according to ISOGG 2017, H2-P96 represents a rare mutation found in contemporary Europe, mostly on Sardinia. A recent study of 1,194 Sardinians (Francalacci et al. 2015) found seven men with the mutation, a frequency of less than one percent. However, ancient DNA from Neolithic sites in Hungary (Haak et al. 2015) and Spain (Günther et al. 2015) suggest that H2-P96 has a much wider distribution in prehistoric Europe. Second, almost all the published data for haplogroup H-M2713 comes from H1a-M69 and its downstream variants, which are commonly found in South Asian populations. Finally, H-M69 represented the main haplogroup H mutation until 2014. Since then the mutation has been placed deeper in the haplogroup H phylogeny, first with H1-M69, and later H1a-M69. For the pre-2014 phylogeny of haplogroup H, the reader is directed to Supplementary 8.2.

Again, H1a-M69 represents almost all of the published haplogroup H data. Additionally, most of the published data come from men living in South Asia, and in particular, Pakistan and India. According to Sengupta et al. (2006), H-M69 attains a frequency of about twenty-six percent among Indians and among the Pakistanis the frequency is about six percent. Elsewhere H-M69 and its variants have been found in some populations of the Middle East, Central Asia and East Asia, where the overall frequency is very low. Additionally, the H-M69 mutation attains a high frequency among several Romani groups in Europe. For further information, the reader is directed to Supplementary Table 8.1, which presents a survey of H-M69 populations.

Sengupta et al. (2006) suggest that H-M69 evolved around thirty thousand years ago. Given the age of this mutation and its moderate frequency among South Asian populations, one finds widespread consensus among the geneticists that identifies H-M69 as an "indigenous" South Asian mutation (e.g. Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008; Debnath et al. 2011; Khurana et al. 2014). Because the phylogeny of H-M69 has been revised, with H-M69 downgraded from a main haplogroup to H1a, the question remains if the new higher level H-M2713 and H1-M3061 mutations are also indigenous South Asian mutations.

Based on the available data as provided above, main haplogroup H-M2713 evolved in Southwest Asia (or the Middle East). As the reader may recall from Chapter 4: Section 2 and the working out-of-Africa hypothesis, the human tribe left Africa about 100 thousand years ago and settled in Southwest Asia. About 50 thousand years ago, with the onset of improved climatic conditions during Marine Isotope Stage 3, part of the human tribe migrated eastwards out of Southwest Asia and colonized South Asia, Australia and East Asia. Part of this expansion included men with the H-M2713 mutation. In South Asia, H1-M3061 diverged from H-M2713. Furthermore, during Marine Isotope Stage 3, part of the human tribe, including those with H-M2713, migrated from Southwest Asia to Europe, the so-called Aurignacian culture. In Europe, H2-P96 diverged from H-M2713.

As noted previously, H-M69 appears to have evolved from H1-M3061 about 30 thousand years ago based on data from Poznik et al. (2016: Supplementary Table 10) and Sengupta et al. (2006: Table 11). Turning now to downstream variants of H1a-M69, three variants are commonly reported in the literature: H1a1-M52, H1a1a-M82, and H1a2a-Apt. Sengupta et al. (2006: Table 11) and Karmin et al. (2015: Table S7) suggest that that H1a1-M52 and H1a1a-M82 evolved during the Neolithic. Dating estimates for H1a2a-Apt, on the other hand, suggest that this mutation evolved during the Mesolithic (Sengupta et al. 2006).

Population reports are inconsistent in reporting internal variation for H-M69. Although H-M69 evolved during the Paleolithic, the available data for its internal phylogeny suggest that the main expansion of this marker occurred during the South Asian Neolithic. Thus the frequency and distribution of H-M69 variation in South Asia may hold important clues for deciphering linguistic diversity in this region. At this point the reader is directed to Supplementary Tables 8.2 and 8.3. Based on these tables, H-M69 appears frequently in South Asia among speakers of languages that fall within the Indo-European, Dravidian, and Austro-Asiatic language families. As previously mentioned in Chapter 7: Section 3, the expansion of Indo-European and Dravidian across South Asia may have a connection with farmers from Southwest Asia who settled at Mehrgarh in the Balochistan region of Pakistan around nine thousand years ago. Over the course of several thousand years, this agricultural trajectory penetrated the Indus Valley of Pakistan and western India, and eventually migrated further eastwards into the Ganges Valley and southwards into southern India and Sri Lanka. Thus the distribution of haplogroup H-M69 variation among Dravidian and Indo-European-speaking populations of South Asia may well be a genetic relic of the Southwest Asian Neolithic package. Austro-Asiatic, on the other hand, appears to have expanded into eastern India about four thousand years ago with the farmers who cultivated a domesticated variety of rice with origins in China (Diamond and Bellwood 2003; Bellwood 2005: 222-227). See, also, Chapter 15: Section 11. Thus haplogroup H-M69 variation may also record the collision of Southwest Asian and East Asian agricultural expansions in eastern India and language shift from Dravidian or Indo-European to Austro-Asiatic.

Section 2. H-M69 and Language Variation in South Asia.

For the purposes of this present discussion, the term "South Asia" presents an overview of linguistic variation in Pakistan and India. Additionally, frequency data for H-M69 is combined with those who have the mutation itself and those who have downstream variants of the mutation: H1a-M52, H1a-M82 and H1a-Apt. This is necessary because many studies have not sequenced H-M69 for informative downstream variants. In the shorthand of the geneticists, I am reporting for H-M69*. Finally, as mentioned previously, the distribution of H1a-M69* reflects demographic processes that began around nine thousand years with the onset of the South Asian Neolithic.

The linguistic diversity found in South Asia is remarkable. In India almost all of the spoken languages fall within one of the four language families: Dravidian, Indo-European, Austro-Asiatic or Sino-Tibetan. In neighboring Pakistan, on the other hand, Austro-Asiatic is absent, but one finds Indo-European, Dravidian, and Sino-Tibetan. With respect to the Indo-

European language family, one main difference between Indian and Pakistani linguistic diversity is that the Indo-European languages of India fall almost exclusively within the Indo-Aryan branch. Hindi, one of India's official languages, and a linguistic heavyweight with over 500 million speakers, provides an example (*Ethnologue* 2018). The Indo-European languages of Pakistan, on the other hand, are a mixture of Iranian and Indo-Aryan. Significant Indo-Aryan languages of this country include Urdu, the official language, and Punjabi, the most widely spoken language (CIA World Fact Book 2018). Within the Iranian branch, Pashto and Balochi are widely spoken.

Turning now to Dravidian languages, eighty-six languages fall within this classification. Brahui is a Dravidian language found in Pakistan. The remaining Dravidian languages, such as Tamil, Telugu and Kannada, are spoken in India. Interestingly, the spatial distribution of Indo-European and Dravidian languages generally follows a geographic pattern in India. Indo-European is found in the North. Dravidian, on the other hand, tends to be present in the south.

Turning now to the Sino-Tibetan language family, the Sino-Tibetan languages of South Asia fall within the Tibetan-Burman branch. Within India, the distribution of Tibeto-Burman languages follows the border that this country shares with Nepal and China. It should be noted that an attempt was made to extrapolate the number of Tibeto-Burman languages spoken in India from the *Ethnologue* website. This proved very difficult, but the figure appears to be around one hundred and twenty-five languages. Examples include Mizo, a language spoken by around 675 thousand people (Ethnologue 2018). The only Sino-Tibetan language listed for Pakistan is Balti, a Tibeto-Burman language with around 327 thousand speakers.

According to Ethnologue (2018), the Austro-Asiatic language family consists of 167 languages. These languages stretch along a geographical expanse than begins in eastern India and end in Malaysia. Within this language family, the Munda branch represents the Austro-Asiatic languages of eastern India. Santhali and Mundari are among the more recognized Munda languages. The Mon-Khmer branch represents East Asian languages. Significant Mon-Khmer languages include Khmer and Vietnamese.

Two studies, Sengupta et al. (2006) and Trivedi et al. (2008), presented frequency data that facilitate analysis of the extent to which haplogroup H-M69 is an informative mutation among the four main language families of India. The problem with the studies is that the sample sizes are very small and as such, ascertainment bias may well be a problem. In order to overcome this problem, Supplementary Tables 8.4 through 8.7 explore the correlation between linguistic and genetic diversity in South Asia by utilizing a large data set of over seven thousand samples gathered from previously published studies. The tables were prepared in order to assess the frequency of H-M69 in South Asians according to language family or language branch. In order to minimize ascertainment bias, the tables excluded data from populations for which the sample size was less than twenty men. The data were then compared against the results obtained by Sengupta et al. (2006) and Trivedi et al. (2008), which are summarized in Supplementary Table 8.8.

Frequency results for H-M69 and its downrange variants are as follows: Indo-European = 17% (Supplementary Table 8.4); Dravidian = 28 % (Supplementary Table 8.5); Austro-Asiatic = 25% (Supplementary Table 8.6); Tibeto-Burman = 6% (Supplementary Table 8.7). It should be noted that the frequency for Tibeto-Burman is probably over-inflated because the H-M69 mutation was not detected in several of the reported Tibeto-Burman-speaking populations, and these populations are not included in the analysis. Indo-European, Dravidian and Austro-Asiatic populations, on the other hand, almost always have the H-M69 mutation. Consequently, the overall frequency data for these three language families are more accurate than that reported for Tibeto-Burman.

The H-M69 frequency data, as just described, supports the two conclusions. H-M69 is a very significant mutation for deciphering the prehistory of South Asian languages. Next, the frequency and distribution of H-M69 in South Asia help in the interpretation of data from other mutations found in the region, especially G-M201, J2-M172, L-M20, T-M184, O-M175, and R1a-Z93. For example, the previous discussion of G-M201 variation in South Asia, as presented in Chapter 7: Section 3, suggested that this haplogroups may well be genetic signature of an expansion of the Indo-European language family from Southwest Asia into South Asia during the Neolithic. Additionally, the same mutation may signal the expansion of the Dravidian from Pakistan to southern India during the same period. Given this interpretation, an interesting question arises. Why does G-M201 decrease in frequency as one moves further east from Pakistan, whereas H-M69 has the opposite clinal pattern, and generally increases in frequency?

The reader is referred once again to Supplementary Table 8.2, which provides a survey of H-M69 variation in Pakistan, and Supplementary Table 8.3, which provides a survey of H-M69 variation in India. Perhaps what is particularly striking about the Pakistani data is that the mutation appears among the Indo-Aryan-speaking Kalash and Iranian-speaking Pathans. The mutation also appears among the Dravidian-speaking Brahui people and among the Burusho, speakers of a language isolate. Of course, the amount of Pakistani data is very small, but nevertheless H-M69 may well have been part of genetic inventory of hunter-gatherers who lived in the Balochistan region when Indo-European-speaking farmers arrived about nine thousand years ago. This suggests that the Neolithic expansion of Austro-Asiatic, Indo-European and Dravidian languages involved complex demographic processes that resulted from the admixture of Southwest Asian farmers and South Asian hunter-gatherers. Taking this a step further, it appears that language shift played a huge role in the evolution of linguistic diversity in South Asia. This is consistent with the available data (see Supplementary Tables 8.4 to 8.7 and Chapter 7: Section 3).

Again, the Neolithic expansion of H-M69 and its variants provide important background information that facilitates analysis of other genetic mutations in South Asia. The ultimate goal of this undertaking is to address several important and longstanding controversies surrounding language variation in the region. One interesting question involves the origins of Dravidian language. Similarly, another question concerns the origins of Austro-Asiatic languages. Another mystery involves Indo-Aryan language and if they arrived in South Asia as the result of a Bronze Age invasion of Central Asian nomads.

Section 3. H-M69 and Language Variation in Central Asia.

H1a-M69 variation in South Asia potentially offers useful data for assessing the purported Central Asian origins of Indo-Iranian languages and the evolution of the Iranian and Indo-Aryan branches. Such a discussion requires the presentation of important background information about the origins of the so-called Central Asian *steppe nomad hypothesis*. It should be noted that at the beginning of the twentieth century, with the discovery of clay tablets at Boğazkale in modern-day Turkey, Hittite became the oldest attested Indo-European language. However, during the nineteenth century Sanskrit was considered the oldest attested Indo-European language. Thus many linguists, such as Max Müller, took a keen interest in this ancient Indo-Aryan language and the Rigveda liturgical texts (for a more detailed discussion see Pedersen 1967 and Arvidsson 2006). From their interpretation (or perhaps misinterpretation) of these texts evolved the idea that the Aryan people were the original speakers of an Indo-European language. During the twentieth century Nazi Germany re-worked the Aryan hypothesis to support their racial and ethnic ideology (for a more detailed discussion see Pringle 2006). The archaeologist Marija Gimbutas then reworked the Aryan hypothesis in a series of articles published between 1952 and 1993. Instead of Aryans she proposed that the first Indo-Europeans were the prehistoric Kurgan people of the Russian steppes. Today the Kurgans have become Central Asian steppe nomads. One most recognized proponents of this current approach to Indo-European origins is the anthropologist David Anthony (for more details see his 2007 monograph). As previously noted in Chapter 7: Section 1, this version of Indo-European origins is controversial. It defies a trend observed elsewhere outside of Indo-European. Worldwide, several large language families co-expanded with early agriculture. Why, then, would Indo-European be an exception to the rule?

Several studies have discussed the Central Asian *steppe nomad hypothesis* and have assessed their potential contribution to the contemporary Indian gene pool (e.g. Kivisild et al. 2003; Cordaux et al. 2004; Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008). Again, this follows the idea that a Bronze Age invasion from Central Asia brought Indo-Aryan languages to South Asia. While final analysis of the Central Asian steppe nomad hypothesis must wait until Chapter 17: Section 9and the discussion of the R1a-Z93 mutation, it is necessary at this time to present H-M69 data that are potentially useful for the discussion.

The H-M69 mutation has been detected in several populations of the Middle East, Central Asia and even East Asia (see Supplementary Table 8.9 for additional information). These populations include those that speak Iranian languages, such as Tajiks in Afghanistan and Tajikistan, as well as Pashtuns in Afghanistan. The mutation has also been detected in Turkic-speaking populations, such as the Uygur of the Xingjian region of China and Uzbeks in Uzbekistan. Thus, these data raise an interesting possibility that prehistoric geneflow between South and Central Asia has been unidirectional, from South to North. Perhaps the geneflow occurred during the Neolithic. The data seem to suggest that the Central Asian Neolithic may have been an expansion of the South Asian Neolithic, which evolved at Mehrgarh in Pakistan roughly nine thousand years ago (see Chapter 5.7, Sections 3 and 4 for additional information). Taking this a step further, perhaps the prehistoric Tocharian people of the Tarim Basin in the Xingjian region and their Indo-European language stand as a linguistic relic of this migration.

Section 4. H-M69 and the Romani Languages.

Haplogroup H has also surfaced as a useful marker for understanding the population history of the Romani people, who are often identified as Roma, and sometimes as Gypsies, a term that is considered derogatory. This population is found throughout Europe. For years scholars have asked whether India is the putative homeland of this group. Language typology certainly points to India as the original homeland of the Romani people. *Ethnologue* classifies the Romani language as part of the Indo-Aryan branch of the Indo-European language family. The historical record also seems to support India as the putative homeland of the Romani (e.g., Tcherenkov and Laederich 2004). Finally, geneticists discovered that haplogroup H-

M69 and its variants are a common Y chromosome mutation among Romani groups in Europe. For example, about seventeen percent of Iberian Romani (Gusmão et al. 2008) and thirty-two percent of Hungarian Romani (Pamjav et al. 2011) have the mutation (see Supplementary Table 8.10 for additional information). A 2012 study published by Rai et al. 2012 analyzed haplogroup H-M69 data that was taken from ten thousand global samples. Based on their analysis, they identified northeast India as the putative homeland of the Romani people.

Section 5. Haplogroup F-M89.

It should be noted that haplogroup H frequencies for South Asia might be underreported in the literature. ISOGG 2017 states that when H-M69 was the main haplogroup mutation, potential H-M2713 and H1-M3061 mutations for South Asia were identified as unspecified variants of haplogroup F-M89 (e.g., Cordaux et al. 2004; Sengupta et al. 2006; Arunkumar et al. 2012; Khurana et al. 2014). Clearly, further testing is needed to attain a more accurate determination of haplogroup H variation in South Asia.

Additionally, it should be explained at this time why we define F-M89 as an evolutionary marker between the main haplogroups and Y-Chromosome Adam rather than a main haplogroup as found in the standard phylogeny (i.e. Karafet et al. 2008). (Note that the M89 mutation is labeled FR-M89 in Supplementary Figure 1.1.) This action was taken because, as demonstrated in the previous paragraph, reported frequency data for F-M89 may well represent more informative haplogroups that had not been identified at the time of publication. Within the global linguistic tapestry, a small number of men probably have the actual F-M89 mutation rather than a variant. However, for purposes of deciphering linguistic variation, the number of men with F-M89 would be too small, and consequently, uninformative.

Section 6. Conclusions for H-M2713.

The discussion of language variation in South Asia continues in Chapter 10: Section 5, and the presentation of J2-M172 variation in this region. Turning now to the present discussion of haplogroup H-M2713, almost all the published frequency data is for H1a-M69 and its downstream variants. H1a-M69 data favor the language-farming hypothesis as an explanation for the origins of Indo-Aryan languages. Furthermore, the same mutation supports the origins of Dravidian languages in Pakistan, and a southwards Neolithic expansion of this language family. Finally, the H1a-M69 places Romani origins among the populations of India.

Chapter 9: Haplogroup I-M170.

Section 1. Introduction.

The reader is directed to Supplementary Figure 1.1 which depicts the important evolutionary steps between Y-Chromosome Adam and the main haplogroups. According to Poznik et al. (2016), haplogroups I-M170 and J-M304 separated from IJ-M429 about 41 thousand years ago. The reader is now directed to Supplementary Figure 9.1 which presents the internal phylogeny of haplogroup I-M170. Within this main haplogroup, I1-M253 and I2-M438 represent the two main internal clades. Both mutations separated from I-M170 about 28 thousand years ago (Underhill et al. 2007). I1-M253 attains a significant frequency among the Germanic and Uralic-speaking populations of Scandinavia. Within I2-M438, three mutations represent the most significant variants: I2a1b-M423, I2a2a-M223, and I2a1a1-M26. Among the populations of the Balkans region of Europe, I2a1b-M423 attains an especially high frequency. Similarly, I2a1a1-M26 attains a significant frequency on the island of Sardinia in the Mediterranean. I2a2a-M223, on the other hand, attains low frequency numbers throughout Europe.

At this point the reader is directed to Supplementary Table 9.1 which provides a survey of haplogroup M170 frequencies across Eurasia. While the haplogroup appears sporadically among some populations in western and central Asia, I-M170 represents the genetic signature of European populations. According to Underhill et al (2007), about twenty percent of European men have the I-M170 mutation. The same study also suggests that I-M170 is the only main Y-chromosome haplogroup that arose on the European continent. The remainder of European Y chromosome variation (e.g. R1b-343, R1a-M420, J2-M172, E1b-V13, G2a-P15, and N1a-Tat) arose from haplogroups that evolved in Asia.

The age of I-M170 (again, 41 thousand years) supports consensus as found among the geneticists. They identify this mutation as the genetic relic of *Homo sapiens* who initially colonized the European continent during Marine Isotope Stage 3 (Sarac et al. 2016; Regueiro et al. 2012; Underhill et al. 2007; Rootsi et al. 2004). Additionally, I-M170 stand as the genetic relic of European populations that survived the last Ice Age. By the beginning of Marine Isotope Stage 2, advancing glacial ice had forced human populations to seek refuge in the southern part of this continent. Beginning about 14 thousand years ago, which roughly corresponds to the beginning of Marine Isotope Stage 1 and the Holocene, warmer weather and retreating ice glaciers allowed human populations to re-colonize the depopulated regions of central and northern Europe.

Section 2. Artifacts, Bones, Climate Change and Ancient DNA.

The initial human colonization of Europe is explained by the paleoclimatological record, human artifacts, skeletal remains, and ancient DNA. During Marine Isotope Stage 3 warmer weather facilitated an expansion of *Homo sapiens* (see Chapter 4: Section 1). This expansion began in Southwest Asia and resulted not only in the initial colonization of Asia (Pope and Terrell 2008), but also Europe (Müller et al. 2011). The timing of this expansion into Europe began with onset of Greenland Interstadial 12, a temporary period of warmer weather during the last Ice Age.

Archaeological for the initial human colonization of Europe comes from Hoffecker (2009) who reports the discovery of artifacts found in Eastern Europe and in Mediterranean region of the continent. These artifacts include spear points and scrappers made from stone. These tools were dated to about 48 thousand years. The artifacts belong to the so-called Aurignacian archeological tradition. Ancient DNA results taken from prehistoric human remains found in Europe Genetic provide genetic support (see Supplementary Table 9.2). As reflected by Reference Sample Nos. 1 through 6 in the table, haplogroups I-M170, C-M130, and NO M214 were part of the genetic inventory (gene pool) of Europe's first Homo sapiens. The distribution of Reference Sample Nos. 1 through 6 further suggests two different migration routes into Europe during Marine Isotope Stage 3: One route encompasses central Europe and the north European coastal plain. The other route follows the Mediterranean Sea. Those that migrated through central Europe probably followed and hunted the large herds of mammals, such as reindeer, mammoths and horses that once roamed the European plain. Those that migrated along the Mediterranean probably exploited marine resources.

During Marine Isotope Stage 2 the weather became colder in Europe. The ice glaciers then reached their maximum southern expansion in Eurasia about 26 thousand years ago, a period that is often referred to as the Last Glacial Maximum (LGM). During the LGM, the ice sheets extended to roughly the fortieth northern parallel in Western Europe (e.g. Gavashelishvili and Tarkhnishvili 2016; Clark 2009). In order to survive, human populations in Europe retreated to refugia in the southern part of the continent.

According to Binney et al. (2016), after the ice sheet had started to contract around 21 thousand years ago, the European landscape above the fortieth northern parallel was a treeless region of tundra. Around fourteen thousand years ago, as the climate warmed, the European tundra also began to contract northwards, leaving behind areas of forest. Around eleven thousand years ago, the tundra reached Scandinavia. Finally, about four thousand years ago, the tundra reached its current location along the Arctic Circle. The contraction of ice and tundra during the Holocene partially explains why the human migrated out of southern European refugia during Holocene. Additionally, it helps to explain how central Europe and Scandinavia were repopulated. Tundra is the preferred habitat for reindeer. As the ice retreated, people migrated northwards to hunt the large mammals, especially the reindeer that roamed the central European tundra. Turning now to Supplementary Table 9.2, Reference Sample Nos. 7, 8, 10, 13, 14, 15, and 16 reflect that haplogroup I-M170 survived the last Ice Age and then expanded into previously depopulated regions after the ice glaciers had retreated. See, also, the discussion in Chapter 17: Section 6.

Section 3. Frequency and Distribution of I-M423.

As previously detailed in Section 1 (above), I1-M253 and I2-M438 evolved roughly 28 thousand years ago. The mutations evolved a point in the prehistory that is defined by the Last Glacial Maximum. Thus, the evolution of both mutations reflects genetic diversification that occurred in southern European Ice Age refugia.

The I2a1b-M423 mutation is a downstream variant of I2-M438. At this point the reader is directed to Supplementary Table 9.3, which provides a survey of population with the I2a1b-M423 mutation. This mutation attains a high frequency among several populations of the Balkans region of Eastern Europe, such as Serbs, Croats, and Bosnians. More moderate frequencies are reported for Albanians, Moldavians, Bulgarians, Ukrainians and Czechs. Genetic studies (e.g. Regueiro et al. 2012; Battaglia et al. 2009) suggest that I2a-M423 evolved roughly eight to nine thousand years ago during the Eastern European Mesolithic. Thus, based on the frequency data alone, it would appear that the I2-M438 upstream mutation evolved in a Balkan refugium, and that the I2a-M423 marker reflects later genetic diversification. See, also, Rootsi et al. 2004.

Two studies (Sarac et al. 2016 and Karachanak et al 2013) explain the evolutionary history of I2a-M423 by defining Paleolithic, Mesolithic and Neolithic components of the contemporary Croatian and Bulgarian gene pools. I2a-M423 represents the Paleolithic component. E1a-V13, R1b-343 and R1a-M420 are potential Mesolithic relics. Neolithic farmers contributed J2-M172 and G2a-P15. Accordingly, the Neolithic transition in Eastern Europe involved a population expansion that originated outside the region and adaptation of a new technology by people already living in the region. Taking this a step further, the demographic effects of the Neolithic transition were probably very complex and unpredictable. Sometimes, the Paleolithic component survived and thrived, as in the case of I-M170. However, C-M130, another relic of the European Paleolithic has disappeared from the contemporary genome of this continent.

Section 4. Frequency and Distribution of I1-M253.

As noted earlier, haplogroup I-M170 bifurcated into two main variants about 28 thousand years ago, I2-M438 and I1-M253. As shown by the Supplementary Table 9.4, I1-M253 attains frequencies between thirty and forty percent in Scandinavia. In this region, the mutation is found among Germanic-speaking Danes, Swedes and Norwegian, as well as Uralic-speaking Finns and Sami. Elsewhere, I-M253 attains more moderate but significant frequencies among the Dutch, Germans, Karelians and Estonians. According to Rootsi et al (2004), after the Last Glacial Maximum I1-M253 expanded northwards from a refugium on the Iberian Peninsula. Ancient DNA data reported by Villalba-Mouco et al. (2019) supports this scenario. The study identified the I1-M253 mutation from Paleolithic remains dated to 13 thousand years ago. These remains come from Balma Guilanyà in Spain.

The reader's attention is now directed to Supplementary Table 9.2 and the survey of ancient DNA results. As shown by the table, some may argue that I1-M253 was not included in the genetic inventory of prehistoric Scandinavia. Four ancient DNA samples were harvested from remains from a burial site in the Östergötland province of southeastern Sweden (see Reference Sample Nos. 13 through 16). According to the researchers, Reference Sample No. 13 belongs to I-M170. Molecular damage, on the other hand, restricts the ability of geneticist to identify the I-M170 variant to which Reference Sample No. 15 belongs. However, it is not I1-M253. Interestingly, Reference Sample Nos. 14 and 16 belong to I2a-M423, the haplogroup I variant discussed in the previous section (3). Again, I2a-M423 attains a heavy frequency in the Balkans. However, it is virtually absent among contemporary Scandinavians.

A salient point for understanding ancient and contemporary genetic variation in Scandinavia emphasizes that the present-day land crossing from central Europe into the region has just one route. It involves a journey through Denmark and across the Öresund Straight via bridge and tunnel into Sweden. However, as explained and illustrated by Sporrong (2003), the landscape in Scandinavian was far different twelve thousand years ago. The water level was around much lower and a larger landmass connected central Europe with Scandinavia. Additionally, the present-day Baltic Sea was a smaller freshwater lake. Consequently, during the early Holocene several different routes presented an opportunity for human settlement in Scandinavia via a land crossing or a short water crossing. The Paleolithic founding populations of Scandinavia probably had I1-M253 and I2a-M423. As suggested by Underhill et al. (2007), I1-M253 entered Denmark via northwestern Europe. Those with I2a-M423 probably entered Scandinavia through another route further east.

The demise of I-M423 and predominance of I-M253 in contemporary Scandinavia is probably the result of demographic and cultural developments that occurred in Scandinavia by the time of the Neolithic transition in this region, roughly five thousand years ago. Agriculture was first adopted in Denmark, and over time the technology spread northwards into the rest of Scandinavia (e.g. Siiräinen 2003). Rapid population growth occurred because agriculture supports higher population density. Thus, a sudden and rapid increase of men with the I-M253 mutation in Denmark, beginning four for five thousand years ago, and their subsequent migration into northern Scandinavia, may well have changed the distribution and frequency of I-M170 variation in the region.

The historical expansions of the Germanic tribes and ethnic Germans may partially explain the presence of I1-M253 in Eastern Europe. Taking this a step further, St. Clair 2014 examined the unexpected frequency of I1-M253 found among several Romani populations in Europe. These populations include Slovak Romani (Petrejčíková et al. 2009), Hungarian Romani (Pamjav et al. 2011) and Iberian Romani (Gusmão et al. 2008). The report suggests that I1-M253 among the European Romani may have been the result of admixture with Crimean Goths.

Section 5. Frequency and Distribution of I2a-M26.

The reader is now directed to Supplementary Table 9.5 which presents the frequency and distribution of the I2a-M26 mutation. On the Mediterranean island of Sardinia, the frequency of this mutation hovers around forty percent. The high frequency of I2a-M26 among Sardinians certainly supports the idea that this island was one of the Ice Age refugia. Additional support for this position stems from the estimated age of the mutation. Morelli et al. (2010) report that I2a-M26 evolved around 18 thousand years ago. Finally, Sondaar et al. (1995) report the discovery of a human phalanx found at the Corbeddu cave on Sardinia, the remains of someone who died about 20 thousand years ago.

The frequency of I2a-M26 ranges from three to sixteen percent among some populations of the Iberian Peninsula. The frequency then drops to around two percent in Ireland. Based on this frequency cline as well as the climatological and archaeological evidence, it appears as though Sardinia was a source population of the Holocene hunter-gatherers that migrated out from the Mediterranean into central and northern Europe. Part of the expansion route from the Mediterranean into central and northern Europe probably involved a journey through the Pyrenees, a mountain range along the contemporary Franco-Spanish border. It should be noted that this remote area of Europe may well have been an Ice Age refugium.

Section 6. Frequency and Distribution of I2a-M223.

The reader may want to examine Supplementary Table 9.6 which provides a survey of populations with the I2a2a-M223 mutation. The frequency and distribution data for the mutation are interesting. Unlike the other three I-M170 variants that have been surveyed (I1-

M253, I2a-M423 and I2a-M26), a frequency cline for I2a-M223 is not observed. I1-M253 frequencies diminish north to south. I2a-M423 and I2a-M26 diminish south to north. I2a-M223, on the other hand, is scattered throughout Europe at frequencies of less than ten percent.

According to Rootsi et al. (2004), I2a-M223 evolved during the early Holocene about 11 thousand years ago. The distribution and frequency patterns of the mutation are such that is difficult to determine exactly where the mutation arose. Nevertheless, the age of the mutation, along with its distribution, suggests that it is a genetic relic of Holocene hunter-gathers that migrated northwards as the tundra contracted. Interestingly, I2a-M223 was detected in remains found at a Neolithic burial site at Atapuerca in Spain. See Reference Sample No. 25 from Supplementary Table 9.2.

Section 7. Significance of I-M170 for Linguists.

7.1. Slavic.

Referring once again to Supplementary Table 9.3, the I2a-M423 mutation appears to be an important marker for deciphering the origin and spread of Slavic languages. Mainstream linguistic opinion that dates the spread of Slavic to historical times, perhaps the fifth century (see Brackney 2007 and discussion in Chapter 17: Section 9). As such, language contact rather than a population expansion explains the contemporary geographic distribution of the language branch.

7.2. The Basque Language Isolate.

Researchers have long suspected that the populations of the Pyrenees are a relic of pre-Neolithic Europe. Partial support for this position stems from the Basque people who speak a language isolate, which is unusual as most Europeans speak an Indo-European language. Since Indo-European languages potentially came to Europe during the Neolithic, roughly eight thousand years ago, one could argue that the Basque languages reflects linguistic diversity from the Mesolithic or earlier.

The expansion of I2a-M26 from Sardinia to the Iberian Peninsula during the early Holocene suggests a pre-Neolithic origin for Pyrenees populations, and with that, the Basque language. Rootsi et al. (2004) reports that about six percent of the Basques have the I2a-M26 mutation. Another report, López-Parra et al. (2009), sampled populations from five remote villages along the Franco-Spanish border and the Pyrenees Mountains region. The frequency of I2a-M26 ranged from three to sixteen percent in the villages. Furthermore, the researchers found C-M130 in two individuals. Given the location of the surveyed population and the fact that C-M130 is rarely found among contemporary Europeans, this was a rather unexpected discovery. This provides additional data which reinforce the idea that the Basque language is a pre-Neolithic relic. As previous outlines in Chapter 6: Section 4, downstream variants of the C-M130 haplogroups are part of founder lineages of Europe, just like variants of haplogroup I-M170. However, I-M170 survived the Last Ice Age whereas C-M130 disappeared from the European gene pool around the beginning of the Holocene (see Chapter 17: Section 3 for additional details).

7.3. Germanic.

Long-standing consensus among linguists places the putative homeland of Germanic languages in Denmark. Based on the frequency data for Scandinavia, haplogroup I-M253 is obviously a significant evolutionary marker for understanding the linguistic prehistory of Germanic languages. St. Clair (2012), a PhD dissertation from the University of California, explored the origins of Germanic languages from a Y-chromosome perspective. One very controversial idea stemming from the dissertation is that Germanic languages evolved as the result of language contact between speakers of proto-Basque, proto-Indo-European and proto-Afro-Asiatic. Perhaps less controversial would be the idea that Germanic languages have considerable time depth. The prehistoric evolution of Germanic partially reflects response to climate change, the isolation of populations from each other, and the Neolithic revolution.

Tundra is the preferred habitat of reindeer. Among prehistoric *Homo sapiens*, reindeer meat was an important source of food (see discussion in Chapter 14: Section 3 and Chapter 17: Section 6 for additional details). Siiräinen (2003) suggests that the human colonization of Scandinavia during the Holocene was facilitated northward contraction of the ice glaciers and tundra. As the tundra contracted, large herds of reindeer migrated out of Central Europe and eventually reached Scandinavia about 12 thousand years ago. Close behind were people that hunted these animals, the so-called Ahrensburg culture, who eventually settled in the region.

Around ten thousand years ago another cultural transition occurred in Scandinavia, the so-called Maglemose people. This cultural transition signaled the arrival of the Holocene, and with that, the landscape transitioned from tundra to forests. This transition forced a change in subsistence strategy. As the result of climate change, people in the region became dependent on marine resources, such as mussels (e.g. Lewis et al. 2016). Inland resources, such as elk, were also an important source of food (e.g. Jessen 2015).

The Ertebølle culture marks the end of the Mesolithic in Scandinavia. As previously detailed in Paper 5.7, Hg. G, Section 2, agriculture expanded across Europe during the Neolithic. The same expansion probably disseminated Indo-European languages throughout the continent. About 7.5 thousand years ago the expansion of the so-called Linear Pottery Culture (LBK) terminated at the Northern European coastal plain. However, agriculture was not embraced by Mesolithic Scandinavians. Rather, the terminal point of the LBK expansion became a cultural boundary that lasted about two thousand years until around 3500 BC and the evolution of the Neolithic Funnel-Beaker culture in Scandinavia.

The reasons for the slow transition to agriculture in Scandinavia remain a mystery. One possible explanation assumes that the Linear Pottery Culture expansion probably carried agriculture though areas of central Europe that was uninhabited or sparsely inhabited by nomadic foragers. In contrast to central Europe, the Mesolithic peoples of contemporary Denmark lived in permanent or semi-permanent settlements. As such the region probably had large population density relative to that of central Europe. Conditions were different in Mesolithic Denmark because of the abundance of marine resources, and with that, the availability of a year-round source of very nutritious food. In other words, the food supply remained stationary and the land supported more people per square kilometer. Taking this a step further, Mesolithic Scandinavians did not need agriculture.

As noted earlier, the Neolithic began in Scandinavia around five thousand years ago. Three different models have surfaced for explaining this transition: human migration, a food shortage, or socio-economic change (Fischer 2002). The idea of human migration deserves particular attention because it undermines or supports the role of language contact theory in shaping the evolution of Germanic languages. Here, genetic data can be an informative for assessing Paleolithic, Mesolithic and Neolithic components of the contemporary Scandinavian gene pool. Karlsson et al. (2006) analyzed almost four hundred DNA samples collected from men in Sweden. With their analysis, they found that the arrival of agriculture in Scandinavia occurred as the result of the adoption of a new technology by people already living in the region rather than an influx of central European farmers. Their conclusion partially follows the heavy frequency of I-M253 mutations in the region, and the low frequency of central European Neolithic markers, especially J2-M172 and G2a-P15. Thus, the Paleolithic component is substantial, and a Neolithic component is minimal.

R1b-343, R1a-M420 and N1a-Tat mutations among modern-day Scandinavians represent potential Mesolithic components. Accordingly, further discussion of language and genes within this region continues in Chapter 14: Section 4.4 and Chapter 17: Section 6. Nevertheless, at that this point it should be noted that the origins of Germanic languages provide valuable insight into the evolution of language variation in Europe. Indigenous pre-Neolithic languages probably shaped the evolution of contemporary European linguistic diversity. Interestingly, linguists have long noted that perhaps a third of the Modern German lexicon lacks an Indo-European cognate (e.g. Vennemann 2000: 241; Waterman 1976:36; Schirmer and Mitzka 1969). Perhaps part of the Germanic lexicon is a relic of the Mesolithic. Furthermore, Mailhammer (2007) suggests that the systematic pattern of ablaut for Germanic strong verbs may have been a featured borrowed from Afro-Asiatic languages. As previously suggested in Chapter 5: Section 5, the E1b-V13 mutation may signal the presence of proto-Afro-Asiatic languages in Mesolithic Europe. However, the mutation is mainly found in the Mediterranean and Balkans. Thus, prehistoric contact between Scandinavia and southeastern Europe seem to be topic worthy of additional research.

Section 8. Conclusions.

I-M170 mutations help to illustrate that language has roots that extend deep into the human prehistory. Language thrives and survives because people have thrived and survived. The evolution of language and I-M170 mutations are relics of successful human adaptation to climate change and the Neolithic revolution in Europe. For example, the evolution of Germanic languages began at the beginning of the Holocene. The landscape transitioned from tundra to forest in Scandinavia, and the reindeer disappeared. People adapted to climate change by harvesting marine resources. This subsistence strategy lasted for those of years. Prehistoric Scandinavians then adopted agriculture. All these factors define the position attained by Germanic within the global tapestry of language variation.

Chapter 10: Haplogroup J-M304.

Section 1. Overview of the J-M304 Main Haplogroup.

The website for the International Society of Genetic Genealogy provides a tree which depicts human Y-chromosome mutations. Anyone who has recently visited the website can surely observe a picture of human genetic variation that has achieved astonishing resolution in the last decade. This evolving picture of human variation is made possible by researchers who have identified thousands of Y-chromosome mutations. Interestingly, the J-M304 mutation was the first Y-chromosome maker that was discovered (Casanova et al. 1985).

At this point the reader is directed to Supplementary Figure 1.1 from Chapter 1. As shown by the figure, haplogroups I-M170 and J-M304 have a close phylogenetic relationship. Both evolved from the IJ-M429 mutation. Poznik et al. (2016: Supplementary Table 10) suggest that the bifurcation of IJ-M429 occurred roughly forty-one thousand years ago. I-M170 probably evolved from IJ-M429 in southeastern Europe (Underhill et al. 2007), and J-M304 evolved from IJ-M429 in Southeast Asia (Semino et al. 2004). At this point the reader is directed to Supplementary Table 10.1 which provides a survey of populations with the J-M304 variation. As shown by the table, J-M304 attains its highest frequency in the Caucasus and in Southwest Asia. Additionally, the table reflects a clinal frequency pattern whereby the frequency of J-M304 steady decreases as one moves further east or west from Southwest Asia.

As the reader may recall from previous discussions (Chapter 5: Section 2; Chapter 7: Sections 2 to 5), hunter-gatherers in Southwest Asia began the transition to agriculture beginning about fourteen thousand years ago. Over time people developed the ability to cultivate crops of legumes and grains. Additionally, the transition to agriculture involved the domestication of sheep and goats. Beginning around nine thousand years ago, some of the Southwest Asian farmers migrated to other regions. This resulted in an expansion of agriculture eastwards into Central and South Asia, northwards into the Caucasus, and westwards into North Africa and Europe. In 1996, Semino et al. published a paper suggesting a good correlation between the distribution of J-M304 variation and the anthropological record. Specifically, the distribution of J-M304 follows the Neolithic expansion of agriculture from the Fertile Crescent of Southwest Asia. Since 1996, other studies have reached the same conclusion (i.e. Arredi et al. 2004; Semino et al 2004, Abu-Amero 2009; Hovhannisyan et al. 2014; Singh et al. 2016).

Clearly the J-M304 haplogroup stands as an important genetic relic of the Southwest Asian Neolithic and its expansion outside the region. The mutation also represents a significant tool for understanding the evolution of several different language groups. In order to harness J-M304 as a tool for linguistic research, researchers must peer into the internal phylogeny of this main haplogroup. The reader is now directed to Supplementary Figure 10.1 which depicts the important downstream variants of the J-M304 main haplogroup. As shown by the figure, J-M304 bifurcates into J1-M267 and J2-M172. Poznik et al. (2016: Supplementary Table 10) suggest that this occurred around thirty-one thousand years ago. As previously noted, this occurred somewhere in Southwest Asia.

Sections 2 to 10 (below) present genetic and linguistic data for several regions of Eurasia as well as North Africa. Indeed, an analysis of the contemporary distribution and frequency of the two main internal variants of the J-M304 main haplogroup reveals that from a regional perspective, J1-M267 stands as an important mutation for deciphering population history in Southwest Asia (Section 2), North Africa (Section 3), and the Caucasus (Section 4). See Supplementary Table 10.2 for additional information. J2-M172, on the other hand, represents a significant marker for Southwest Asia (Section 2), the Caucasus (Section 4), South Asia (Section 5). Central Asia (Section 7), and the Mediterranean region of Europe (Section 9). See Supplementary Table 10.3 for additional information. When analyzed from a linguistic perspective (see Supplementary Table 10.4), J1-M267 appears to be a significant marker for Afro-Asiatic-speaking populations in Southwest Asia and North Africa (Sections 2 and 3). Additionally, the marker attains a high frequency among North Caucasian-speaking in Caucasus region (Section 4). Turning now to a linguistic analysis of J2-M172 (see Supplementary Table 10.5), this marker represents a potential tool for deciphering the population history of Indo-European-speaking populations in several regions, including the Caucasus (Section 4), Iran (Section 6), Central Asia (Section 7), East Asia (Section 8), and Europe (Sections 9 and 10). Additionally, J2-M172 represents a significant marker among Indo-European and Dravidian-speaking populations of South Asia (Section 5).

In published studies, geneticists occasionally report a "demic diffusion" of genes as a result of the Southwest Asian agricultural expansion. The term describes a demographic model that is poorly understood. Moreover, it fails to accurately capture the expansion as carried by contemporary genetic and archaeological perspectives. A discussion of this topic is presented in Section 11 as it helps to clarify the co-migration of languages and genes during Neolithic, information that is potentially useful for linguists. Finally, Section 12 emphasizes that several Y-chromosome haplogroups support the *early farming dispersal hypothesis*. As such, the hypothesis presents a robust model of language evolution.

Section 2. Linguistic and J-M304 Variation in Southwest Asia *(Except Iran)*.

Southwest Asia not only represents one of several regions in the world where agriculture evolved, but also the potential homeland where Indo-European and Afro-Asiatic languages evolved (see discussions in Chapter 5: Section 3, Chapter 7: Sections 2 and 3). Two published reports (Chiaroni et al. 2008; Chiaroni et al. 2010) found that in Southwest Asia a statistically significant correlation exists between annual precipitation and the evolution of sedentary dry land agriculture as well as pastoralism. In other words, people grow crops in Southwest Asia where rainfall is abundant. Where rainfall is less than four hundred millimeters per year, people in Southwest Asia tend to herd goats and sheep. Taking this a step further, an interesting pattern emerges whereby people that utilize pastoralism speak Afro-Asiatic languages, and those that grow crops tend to speak Indo-European languages.

The evolution of language in Southwest Asia potentially represents a population split during the Neolithic, and with that, agricultural specialization in the region. Platt et al. (2017), based on a synthesis of the genetic, archaeological and climate data, suggest that the split occurred roughly eight to nine thousand years ago in eastern Turkey or perhaps the southern Caucasus. Balanovsky et al. (2017b), based on an analysis of the genetic data, suggest that topography shaped J1-M267 and J2-M172 variation in Southwest Asia. From their study, one could argue that the Anatolian, Armenian, Iranian and Mesopotamian plateaus served as corridors that facilitated an expansion of farming, language and Haplogroup J-M304 mutations from the Black Sea during the Neolithic.

Based on the distribution of J2-M172 variation, especially J2a-M47 (Supplementary Table 10.12), J2a-M67 (Supplementary Table 10.7), J2a-L24 (Supplementary Table 10.6) and

J2b-M241 (Supplementary Table 10.8), both genes and farming must have dispersed very rapidly from the Black Sea during the Neolithic. The dispersal pattern seems almost akin to the remains of a supernova explosion. For example, the J2b-M241 variant is detected not only in several populations in eastern India, but also in Mediterranean Europe, and even in Western Europe among the Flemish people. A catastrophic flood of the Black Sea may well explain this dispersal pattern. At the end of the last Ice Age, the Black Sea was a freshwater lake. According to Ryan and others (2003), about 8,500 years ago an earthen dam collapsed at the Straights of Bosporus due to melting glacial ice and the corresponding rise of sea level in Mediterranean Sea. A "catastrophic flooding of the Black Sea" occurred creating the current saltwater sea which is much larger than the original fresh water lake. Interestingly, Ryan and others in their 1997 discussion of the Black Sea flood suggested that this event resulted in a dispersal of farmers towards Europe, accelerating the Neolithic transition on this continent (see, also, Karachanak et al. 2013). Taking this a step further, it seems that the flood not only affected genetic variation in Europe, but also in North Africa and South Asia.

Section 3. Linguistic and J-M304 Variation in North Africa.

At this point the reader is invited to examine the regional data for North Africa as reported in Supplementary Tables 10.2 and 10.3. The frequency of J2-M172 is low throughout the region. J1-M267, on the other hand, attains moderate frequency levels. Interestingly, the distribution of J1-M267 in the Levant region of Southwest Asia as well as in North Africa represents a dispute among the geneticists. Some have taken the position that J1-M267 variation was considerably shaped by the historical spread of Islam from the Arabian Peninsula (i.e. Semino et al. 2004; Capelli et al. 2006; Zalloua et al. 2008; El Sibai et al. 2009; Triki-Fendri et al. 2015). Other report that J1-M267 represents much earlier agricultural expansions during the Neolithic (Arredi et al. 2004; Abu-Amero et al. 2009; Tofanelli et al. 2009b; Fadhlaoui-Zid et al. 2011a; Fadhlaoui-Zid et al. 2013).

Those that favor a Neolithic expansion of J1-M267 seem to have the historical record on their side, which presents little if any evidence to associate the spread of Islam with mass migration. This, in turn, raises a linguistic argument in favor of the J1-M267 Neolithic hypothesis. Taking this a step further, Berber languages in North Africa represent *in situ* diversification of a Proto-Afro-Asiatic language that swept across the region during the Neolithic (see Chapter 5: Sections 3 and 4). Similar arguments could be made for Omotic, Cushitic and Egyptian. Arabic, on the other hand, represents *in situ* diversification of Proto-Afro-Asiatic-speaking people that remained in Southwest Asia during the Neolithic.

The J1a-P58 mutation, a downstream variant of J1-M267, also favors the J1-M267 Neolithic hypothesis. As mention in the previous section (2), climatological and genetic data suggest that around nine thousand years ago two different populations arose in the Middle East. One population adopted dry land agriculture as a subsistence strategy. The genetic signature of dry land agriculture is J2-M172, and by extension, J2-M172 becomes the genetic signature of Indo-European languages. The other population adopted pastoralism. J1-M267 represents their genetic signature, and by extension, this mutation became the genetic signature of Afro-Asiatic languages. These conclusions stem from two different published reports. One of the reports (Chiaroni et al. 2010) focuses specifically on the J1a-P58 mutation. The study identifies this downstream variant of J1-M267 as a particularly informative mutation for explaining the Neolithic expansion of pastoralism and Afro-Asiatic languages in Southwest Asia and North Africa. According Chiaroni et al. (2010), J1a-P58 evolved roughly nine thousand years ago within the Taurus and Zagros mountains of eastern

Turkey. Thus a catastrophic flooding of the Black Sea could have accelerated a rapid southward dispersal of genes and farming. It appears as though the Mesopotamian Plateau facilitated this dispersal.

At this point the reader is invited to examine Supplementary Table 10.9, which sorts J1a-P58 variation according to language, and Supplementary Table 10.10, which sorts J1-P58 variation according to region. The amount of J1-P58 data is rather limited because J1-M267 is often not sequenced for informative downstream mutations. Nevertheless, the available data point to Southwest Asia as the region where the mutation evolved. Additionally, the same data provided additional evidence that the mutation co-expanded across North Africa with Afro-Asiatic languages during the Neolithic.

Section 4. Linguistic and J-M304 Variation in the Caucasus.

For the purposes of this discussion, the Black Sea defines the western boundary of the Caucasus region. The eastern boundary is defined by the Caspian Sea. The 40th and 44th parallels roughly define the southern and northern boundaries. The region includes parts of Russia, as well as Armenia, Georgia, Azerbaijan, and arguably the Armenian plateau in Turkey. In a previous discussion of the Neolithic transition in the Caucasus it was noted the Southwest Asian agricultural package arrived in the Caucasus region about 8.5 thousand years ago (see Chapter 7: Section 5). Farmers with the J1-M267 and J2-M172 mutations were part of this expansion that was probably facilitated by the Armenian Plateau. Geographical isolation and endogamy then shaped the frequency patterns of both mutations in the Caucasus region. This conclusion stems from a synthesis of data from several published reports (Balanovsky et al. 2011, Herrera et al. 2012; Yunusbayev et al. 2012; Hovhannisyan et al. 2014; Karafet et al. 2016; Balanovsky et al. 2017b).

Within the Caucasus region, four different language families are found: Indo-European, Kartvelian, North Caucasian, and Turkic. The frequency pattern of J1-M267 and J2-M172, as found among the various linguistic groups of the region, is rather interesting. J1-M267 attains a heavy frequency among North Caucasian speakers, and a moderate frequency among some Armenian-speaking populations (see Supplementary Tables 10.2 and 10.4). As shown by Supplementary Tables 10.3 and 10.5, low to moderate frequencies of J2-M172 are found among North Caucasian-speaking population, whereas the frequency of J2-M172 ranges from moderate to heavy among the Armenians. Among Turkic-speaking population, J2-M172 exhibits a moderate frequency. A similar frequency level of J2-M172 is seen for Iranian and Kartvelian-speaking populations. These data support the position that Kartvelian and North Caucasian are ingenious languages of the Caucasus, whereas Turkic and Indo-European were “imported.” Taking this a step further, language maintenance and language shift shaped linguistic variation in the Caucasus (see, also, Chapter 7: Section 5).

Both the genetic and linguistic data from the Caucasus help to pinpoint the geographic homeland of Indo-European languages to an area near the southern shore of the Black Sea. As such, this supports and further expands the argument made earlier (Section 2). Specifically, a catastrophic flood accelerated a Neolithic dispersal of farmers, genes and languages. It should be emphasized that among the extant branches of the Indo-European language family, Armenian-speaking populations are geographically closest to the putative Indo-European homeland. Among the extinct Indo-European languages, Hittite is closest.

Section 5. Linguistic and J-M304 Variation in South Asia.

For the purposes of this discussion, South Asia consists of contemporary Pakistan and India. J1-M267 does not appear to be a significant mutation among the populations of this region (see Supplementary Table 10.2). On the other hand, J2-M172 represents an extremely important for deciphering the genetic history of these populations, especially those that speak Indo-Aryan and Dravidian languages. For both groups, the frequency of the mutation ranges from low to moderate (see Supplementary Table 10.5 and data for Dravidian and Indo-European).

The Neolithic transition in South Asia was previously introduced in Chapter 7: Section 3. As detailed in this section, the Neolithic transition has African, East Asian and Southwest Asian components. Expanding now upon the Southwest Asian component, a synthesis of the geological, genetic and linguistic data suggests that around nine thousand years ago a catastrophic flooding of the Black Sea (see Section 2) forced Indo-European-speaking farmers to migrate eastwards via the Iranian plateau. Shortly thereafter, they settled in the Balochistan region of Pakistan. At this point Dravidian-speaking farmers adopted agriculture from Indo-European-speaking farmers. Within a short period of time, perhaps within 500 years, Dravidian and Indo-European-speaking farmers then expanded into the Indus Valley of western India. Dravidian-speaking farmers then migrated southwards and by around five thousand years ago they had settled in southern India and Sri Lanka. Indo-European-speaking farmers, on the other hand, migrated eastwards from the Indus Valley. By around five thousand years ago, they had settled in the Ganges Valley.

Based on the migration scenario discussed in the previous paragraph and linguistic relationships within the Indo-European language, it is plausible Indo-Iranian languages evolved in Balochistan. Taking this a step further, when Indo-Iranian-speaking farmers migrated into the Indus Valley, Indo-Iranian diversified into the Indo-Aryan branch. Iranian languages, on the other hand, represent diversification of Indo-Iranian among those that remained in Balochistan. This model of Iranian and Indo-Aryan origins defies, of course, an alternative long-standing model of Indo-Aryan and Iranian language origins. According to this model, a culture called the "Aryans" had conquered India, perhaps during the Bronze Age, and imposed their languages on the people of the region. As noted previously in Chapter 8: Section 3, several Y-chromosome studies have rejected this language model. Moreover, in 2016 a study appeared (Singh et al.) which provides a substantial amount of J2-M172 data for South Asia. These data, in turn, provide additional support for the position that the Southwest Asian agricultural package and Indo-European languages co-migrated to South Asia during the Neolithic. By extension, additional support is generated for the *early farming dispersal hypothesis*, the idea that many of the language families in the world co-expanded with early agriculture.

In order to discuss the study published by Singh et al in 2016, it is necessary to explore, once again, the internal phylogeny of the J2-M172 mutation. As shown by Supplementary Figure 10.1, J2-M172 bifurcates into J2a-M410 and J2b-M12/M102. Within the J2b-M12/M102 clade, most of the South Asian variation falls within J2b2a-M241. It should be emphasized at this point that the J2b-M241 mutation is found not only in South Asian populations, but also in population to the east, across a vast geographical expanse that extends to Western Europe. This pattern of J2b-M241 variation supports, in turn, the idea that between eight and nine thousand years ago a catastrophic flood occurred in the vicinity of the Black Sea. Some fled to South Asia, and some fled to Europe.

Is should be noted that researchers have been unable to further resolve J2a-M410 variation in South Asia. Singh et al. (2016) propose the Z2396 mutation. However, the location of the mutation within the J2a-M410 phylogeny still has not been determined. Thus, Singh et al (2016) caution that better resolution of the J2a-M410 may alter contemporary opinion that identifies J2-M172 as a Neolithic component of the South Asian gene pool, and as such, this would undermine the *early farming dispersal hypothesis*. However, a change in opinion seems unlikely. Overwhelming support for the hypothesis comes from the fields of genetics, archaeology, and linguistics.

Focusing now on the study published by Singh et al. (2016), they suggest that the *early farming dispersal hypothesis* is undermined by a lack of G2a-P15, E1b-V13 and R1a-M269 mutations in South Asia. As discussed in Paper 5.17, Section 6, the R1b-M269 mutation is marker of the Western European Mesolithic. Concerning G2a-P15 variation in South Asia, it is difficult to explore the role that this mutation played in the South Asian Neolithic because most of the published data for region has failed to sequence downstream variants of the G-M201 main haplogroup. Nevertheless, Chapter 7: Section 3 makes a compelling argument that identifies G-M201 as an informative marker for South Asia, and that this marker supports the *early farming dispersal hypothesis*. Additionally, the same section discusses the G2b-M377 mutation and asserts that this is an especially strong marker for exploring the South Asian Neolithic. The discussion includes Supplementary Table 7.9 from Chapter 5 which demonstrates that G2b-M377 has been detected across a vast geographical expanse, from South Asia to Western Europe. This suggests a similar dispersal pattern as just described for the J2b-M241 (see Supplementary Table 10.8). Finally, it should be emphasized that the E1b-V13, like R1b-M269, is a European Mesolithic marker (see Chapter 5: Section 5) and as such, is *not* an informative marker for exploring Neolithic expansions.

Section 6. Linguistic and J-M304 Variation in Iran.

In Section 2, Iran was excluded from the discussion of linguistic and J-M304 variation in Southwest Asia. The reason for this decision stems partly from the fact that Iranians speak an Indo-European language while the rest of region speaks Afro-Asiatic. Additionally, among speakers of Iranian languages the frequency of J1-M267 is generally low, whereas the frequency of J2-M172 is moderate (cf. Supplementary Tables 10.2 and 10.3 and data for Southwest Asia). An explanation follows early agricultural dispersals from the Black Sea roughly nine thousand years ago. Some dispersed southwards onto the Mesopotamian plateau, which served as a natural corridor that directed expansions into Syria, Iraq and the Arabian Peninsula. Relics of this dispersal included elevated frequencies of the J1-M267 mutation as well as Afro-Asiatic languages and pastoralism. However, a strong component of contemporary Iranian and South Asian gene pools evolved from an early farming dispersal from the Black Sea via the Iranian Plateau. Relics of this dispersal include elevated frequencies of the J2-M172 mutation, sedentary crop agriculture, the evolution of the Indo-Iranian branch from Indo-European language family, and the subsequent evolution of Indo-Aryan and Iranian branches from Indo-Iranian (see Section 5 above for additional details).

Section 7. Linguistic and J-M304 Variation in Central Asia.

For the purposes of this discussion, Central Asia is defined as Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. The Central Asian

Neolithic was previously introduced in Chapter 6: Section 7.3 and Chapter 7: Section 4. The Neolithic transition in Central Asia has an indigenous component as well as a Southwest Asian component. The indigenous component stems from the domestication of the horse, which occurred in north-central Kazakhstan about 5.5 thousand years ago. The Southwest Asian agricultural package, on the other hand, arrived in Turkmenistan about eight thousand years ago. This package consisted of both crops and herd animals. Then, over the course of about four thousand years, elements of the Southwest Asian agriculture package advanced into Tajikistan, Uzbekistan and eventually Kazakhstan. This model of the Central Asian Neolithic stems from archeological data (see Chapter 6: Section 7.3 and Chapter 7: Section 4.). Furthermore, it agrees with the genetic data. See Supplementary Table 10.3 and data for Turkmen, Uzbeks and Tajiks, Pashtuns and Kazakhs.

Iranian and Turkic languages are clearly part of the Central Asian linguistic tapestry. An effort to identify the source of Iranian languages for this region represents important evidence that either undermines or supports the *steppe nomad hypothesis* of Indo-European origins (see Chapter 7: Section 1 for additional details). This hypothesis places the evolution of Indo-Iranian languages somewhere in the western part of the Eurasian steppes, and more specifically, north of the Caspian Sea in Russia. The hypothesis then suggests that steppe nomads advanced southwards and invaded modern-day Iran, Pakistan and India during the Bronze Age. Following the invasion, according to the hypothesis, Indo-Iranian further evolved into the Iranian and Indo-Aryan branches. However, as detailed previously in Section 5, the archeological record and genetic data suggest that Indo-Iranian and Iranian languages appear to have evolved in the Balochistan region of Pakistan, and that Indo-Aryan evolved in the Indus Valley of India. Thus, the problem with the *steppe nomad hypothesis* is that it places the evolution of Indo-Iranian languages in the wrong region. Rather, the *early farming dispersal hypothesis* stands as a more robust model for explaining the evolution of Indo-Iranian.

In 2012 Frachetti published an interesting paper that challenges traditional approaches to Central Asian anthropology. In order to discuss the traditional view, it should be noted that Central Asia lies in the eastern part of the Eurasian steppes. Traditional approaches to anthropology in the Central Asia are deeply rooted in the *steppe nomad hypothesis*, which posits conquest of the region by nomadic horse mounted invaders from the west. Frachetti (2012), on the hand, suggests that numerous cultures evolved *in situ* along the vast Eurasian steppes. He then argues that the conquest of Central Asia by these steppe nomads from the west seems inconsistent with the archeological record. A key component of Frachetti's argument stems from the archaeological record. According to this data, the horse was initially domesticated in north-central Kazakhstan. Trade networks, rather than invasion and conquest, as posited by the *steppe nomad hypothesis*, later extended the use of domesticated horses throughout Eurasia. He reports that one these networks may well have been the Inner Asian Mountain Corridor.

The Inner Asian Mountain Corridor extends from the Hindu Kush Mountains of Pakistan and runs in a northeasterly direction to the Altai Mountains of Siberia. This route facilitates travel through and past the Pamir, Tian Shan and Dzhungar Mountains. Interestingly, Frachetti (2012) and Spengler et al. (2014) suggest that this corridor facilitated dispersal of the Southwest Asian agricultural package into Central Asia. Taking this a step further, the same corridor may well have facilitated the spread of Iranian languages from Pakistan into Afghanistan and the rest of Central Asia. Contemporary linguistic relics of this expansion may include Pashtun and Tajik.

Section 8. Linguistic and J-M304 Variation in East Asia.

Tocharian represents an extinct branch of the Indo-European language family. It was spoken until about a thousand years ago in the Tarim Basin and Xinjiang region of eastern China. Previously, in Chapter 8: Section 3, it was reported that researchers found the H-M69 mutation among some populations of this region. Similarly, the researchers have detected the J2-M172 mutation among Turkic-speaking Uzbeks and Uyghur, and Iranian-speaking Tajiks living in the Xinjiang region (see Supplementary Table 10.3 and data for East Asia). This discovery, along with the presence of H-M69 in the region, may well represent a prehistoric genetic relic of geneflow between South Asia and East Asia via the Inner Asian Mountain Corridor (see Section 7). This migration may well have carried Indo-European languages into the Tarim Basin, and as such, may explain the mysterious origins of Tocharian.

Section 9. Linguistic and J-M304 Variation in the Mediterranean Region of Europe.

For the purposes of this discussion, Albania, mainland Greece, and mainland Italy define the Mediterranean region of Europe. Additionally, the region consists of Cyprus, the numerous Greek islands (e.g. Crete), as well as Sicily, Sardinia and Corsica. Within the region, the frequency of J1-M267 is low (see Supplementary Table 10.2 and data for the Mediterranean). However, the J2-M172 mutation attains moderate frequencies (see Supplementary Table 10.3 and Mediterranean region).

As previously detailed in Chapter 7: Section 2, Indo-European languages and farming co-expanded across Europe during the Neolithic. This expansion followed three different trajectories. One trajectory follows the maritime colonization of the Mediterranean islands. One genetic relic of this expansion appears to be the J2a-M319 mutation (see Supplementary Table 10.11). The second trajectory follows the expansion of farming along the southern Mediterranean coast of mainland Europe. A particularly strong genetic relic of this expansion is J2a-M67 (see Supplementary Table 10.7). Accordingly, from a Y-chromosome perspective, the evolutionary history of the Albanian, Greek and Italic branches of the Indo-European language potentially includes *in situ* diversification of Neolithic farming languages.

Section 10. Linguistic and J-M304 Variation in Eastern Europe, Western Europe and Scandinavia.

Another Neolithic trajectory for Europe follows the expansion of farming from southwestern Asia into the Balkans and later into Central Europe and Scandinavia. In Scandinavia the J-M304 mutation is found in about four percent of Swedes (Karlsson et al. 2006). Further north in Finland, J-M304 was not detected in a study that sequenced over five hundred samples (Lappalainen et al. 2008). These data for Scandinavia follows a pattern throughout Europe whereby the frequency of J2-M172 steadily decreases from Albania, either from east to west, or south to north. Several studies provide an explanation (e.g. Capelli et al. 2007; King et al. 2008; Battaglia et al. 2009; Karachanak et al. 2013). They suggest that the Neolithic spread of farming from Southwest Asia to Europe was carried by farmers with the J2-M172 and G-M201 mutations. When they migrated to Europe, the continent was inhabited by hunter-gatherers. The gene pool of the hunter-gatherers included E-V13 (Chapter 5), variants of the R-M207 main haplogroup Chapter 17), and variants of the I-M170 main haplogroup (Chapter 9). In some cases, new populations were formed through admixture between farmers and hunter-gatherer. Alternatively, hunter-gatherer simply adopted farming.

These two demographic scenarios, along with rapid population growth that agriculture facilitates, eventually produced populations in Europe with varying frequencies of genes of Neolithic, Mesolithic or Paleolithic origin. Although the Neolithic farming genes of Southwest Asia eventually disappear among Northern European populations, the Indo-European languages they spoke thrived and survived, with not only Albanian, Greek and Italic as the linguistic relics, but also Slavic, Celtic and Germanic.

Section 11. Modeling the expansion of the Southwest Asian Neolithic.

The term "demic diffusion" requires clarification at this point as it seems to have caused much confusion among researchers. The previous section (10) suggests that European Neolithic involved admixture and acculturation. This also occurred in other regions, such as Central and South Asia. In other words, about 8.5 thousand years ago, farmers with Haplogroup J-M304 fled the Black Sea in all directions with their crops and herd animals. As they expanded, farmers with Haplogroup J-M304 variants encountered hunter-gatherers with other Y-chromosome mutations, such as R1b-M269 in Europe, or H-M69 in India. Sometimes, both Neolithic farmers and hunter-gatherers formed a new population that utilized agriculture. This, in turn, would have decreased the frequency of Haplogroup J-M304 variation. Alternatively, hunter-gatherers outside of Southwest Asia simply adopted the new technology through cultural contact with farmers. This, also, would have decreased the frequency of J-M304 variation.

The term "demic diffusion" often appears in genetic studies that discuss the expansion of the Southwest Asian agricultural package (e.g. Balaresque et al. 2010; Regueiro et al. 2012; Singh et al 2016). Development of the demic diffusion model is often attributed to interdisciplinary collaboration between the archaeologist Albert Ammerman and the geneticist Luca Cavalli-Sforza (e.g. Ammerman and Cavalli-Sforza 1984). Both researchers proposed that the expansion of the Southwest Asian agricultural package involved a migration of a small number of farmers into previously *uninhabited* areas. A population explosion followed their arrival because agriculture is a subsistence strategy that potentially supports high population density with a given region.

It seems as though some researchers have failed to understand that the demic diffusion model requires an agricultural expansion into an *uninhabited* area. While the demic diffusion model may explain the Neolithic transition in some regions, such as North Africa, the term also surfaces when admixture and acculturation provide models that are more consistent with the archeological and genetic evidence, such as in Europe or South Asia. For linguists, the salient point here that is language shift among hunter-gatherers stands a component of the Neolithic transition in many regions of the world.

Section 12. Important Phylogenetic Relationships.

As detailed in Sections 2 and 3 (above), the J1-M267 mutation is a valuable marker for exploring the evolution of Afro-Asiatic languages. Similarly, J2-M172 is a valuable marker for exploring the evolution of Indo-European languages (see Section 4 to 10). Moreover, other Y-chromosome markers carry a similar story of how both language families evolved. Specifically, these markers are variants of the G-M201, E-M96, and H-M2713 main haplogroups. Accordingly, a discussion of important phylogenetic relationships will showcase the following: the *early farming dispersal hypothesis* represents a robust model of language evolution.

One interesting observation is that wherever you find “G,” you find “J.” For example, the distribution pattern of G-M201, along with that of J2-M172, suggests that Indo-Aryan and Dravidian languages arose in eastern Pakistan (see Chapter 7: Sections 3 and 5). Furthermore, both G-M201 and J2-M172 reflect unidirectional gene flow from South Asia into Central Asia, and as such, this undermines the *steppe nomad hypothesis*. Rather, these data support the *early farming dispersal hypothesis* and the position that Indo-Iranian probably evolved in present-day Pakistan (Chapter 7: Section 4, and this current paper, Sections 5 and 7). Finally, G-M201 (especially G2-L91 and G2-L497) and J2-M172 variants (especially J2a-M67) reflect a co-expansion of agriculture and Indo-European languages across Europe during the Neolithic (Chapter 7: Section 2; and Chapter 10: Sections 9 and 10; and Supplementary Table 10.7).

Another interesting observation is that G-M201 and J2-M172 variation reflect rapid population growth following the adoption of agriculture in Southwest Asia. Then, a catastrophic flood produced a supernova-like expansion of genes, farmers and languages from the Black Sea about eight or nine thousand years ago. This is consistent with the atypical distribution of G-M201 and J2-M172 variants. Deeply rooted mutations within the Y-chromosome phylogeny generally reflect localized *in situ* diversification of upstream mutations. However, downstream variants rooted deep within the G-M201 and J2-M172 phylogeny are spread across a vast geographical expanse. For example, the G2-L30 variant is found in Judeo Tats, Bagvalal, and Nogais of the Caucasus region (Karafet et al. 2016). The same mutation is also detected in Flanders (Larmuseau et al. 2014). Similarly, J2b-M241 is found both in eastern India and in Flanders (Larmuseau et al 2014, Singh et al. 2016).

Focusing now on H-M69 and J2-M172, both have been detected in the vicinity of the Tarim Basin of eastern China (Paper 5.8, Hg. G, Section 2 and this current paper, Section 8). This discovery is unusual as both mutations only represent a small fraction of the East Asian gene pool (Zhong et al. 2011). As such, their presence in the region certainly seems unusual. They probably arrived in the region as the result of gene flow from South Asia to Altai Mountains via the Inner Asian Mountain Corridor (see Section 7). This migration through the corridor probably carried Indo-European language from South Asia into Central and East Asia.

Finally, like Indo-European, Afro-Asiatic also arose in Southwest Asia (see, also, Chapter 5: Section 3). E-M81, E-M34 and E-V22 represent expansion of agriculture and Afro-Asiatic out of Southwest Asia into Africa. E-M81 expanded in North Africa (Chapter 5: Section 4). E-M34 and E-V22 expanded into North Africa and East Africa (Chapter 5: Sections 4 and 5). J1-M267 (especially J1a-P58) co-expanded with E-M81, E-M34 and E-V22 (see Sections 2 and 3 of this paper).

Section 13. Conclusions.

Topography also appears to be one of several factors that explain prehistoric language expansions. Both J1-M267 and J2-M172 evolved in Southwest Asia roughly 30 thousand years ago. A population having both mutations appears to have split following a catastrophic flood at the Black-Sea roughly eight to nine thousand years ago. J1-M267 correlates with well with farmers that fled the Black Sea via the Mesopotamian Plateau. From this dispersal evolved the pastoral food economies in Southwest Asia and North Africa. Additionally, the dispersal correlates well with the evolution of Afro-Asiatic languages. J2-M172, on the other hand, correlates well with farmers that fled the Black Sea utilizing one of three different routes: the Anatolian Plateau, the Armenian or the Iranian Plateau. This dispersal resulted in an expansion of the Southwest Asian Neolithic to several different regions. These regions include Europe, Iran, South Asia and Central Asia. The dispersal also correlates well with the expansion of Indo-European languages. Dispersal via the Iranian Plateau helps to explain the evolution of Indo-Iranian and Dravidian languages. Language variation in the Caucasus is partially explained by the dispersal through the Armenian plateau. Finally, dispersal via the Anatolian plateau helps to decipher the evolutionary history of Indo-European languages on the European continent.

Chapter 11: Haplogroups L-M20 and T-M184.

Section 1. Overview of LT-L298 Variation.

At this point the reader is invited to review Supplementary Figure 1.1 from the first chapter. As shown by the figure, the LR-M9 mutation bifurcates into LT-L298 and KR-M526. L-M20 and T-M184 then diverge from LT-L298. According to Poznik et al. (2016: Supplementary Table 10), LT-L298 evolved about 44 thousand years ago. Additionally, the same study suggests the divergence of haplogroups L-M20 and T-M184 from LT-L298 occurred about 41 thousand years ago.

Historically, researchers have experienced difficulty in identifying the position occupied by L-M20 and T-M70 within the overall Y-chromosome phylogeny. As previously detailed in Chapter 1, the first Y-chromosome mutation was identified in 1985. By 2002 advances in sequencing technology allowed researchers to identify over two hundred Y chromosome markers. However, at this point geneticists were utilizing at least seven different nomenclature systems to label these mutations. This hindered the potential of the Y-chromosome as a research tool. Standardization was clearly needed and that year the Y-Chromosome Commission (YCC 2002) issued what is still the standard nomenclature for Y-chromosome haplogroups.

In the YCC 2002 report, L-M20 and K2-M70 both appeared downrange from K-M9. Karafet et al. (2008) then re-labeled K2-M70 as T-M70 and placed this mutation along with L-M20 downstream from K-M9. Chiaroni et al. (2009) later identified the M526 mutation as a downstream variant of M9. Shortly thereafter, Mendez et al (2011) reported the discovery of the M184 mutation, which became main haplogroup T-M184. The M70 mutation, in turn, became T1-M70. The same study also identified LT-L298 as a sister clade of M526. Finally, the study reported that LT-L298 unites T-M184 and L-M20. In 2012, the International Society of Genetic Genealogy (ISOGG) repositioned the M70 mutation within the Y-chromosome phylogeny, and T1-M70 became T1a-M70.

The above discussion of the L-M20, T-M184 and T1a-M70 mutations is provided to emphasize three main points. First, LR-M9 and its downstream mutations have been difficult to position within the Y-chromosome phylogeny. This is an important point that will resurface in Chapter 12 and the discussion of unclassified LR-M9 mutations. Secondly, researchers should know that data for T-M184 must be gleaned from pre-2011 studies that report data for K2-M70 and T-M70. Finally, Section 2 (below) presents arguments that favor the treatment of LT-L298 as a main haplogroup within the Y-chromosome phylogeny. As such, L-M20 and T-M184 should be viewed as its two main downstream variants.

Section 2. Phylogenetic Relationships.

The reader is asked to note that L-M20 and T-M184 are comparatively rare mutations. They appear sporadically among the populations in several different regions of Eurasia, as well as in North and Sub-Saharan Africa. Furthermore, when detected, L-M20 and T-M184 generally attain a frequency of less ten percent. Nevertheless, when viewed as subclades of a new LT-L298 mutation, they become an important component in understanding the correlation between genetic and linguistic diversity. Accordingly, this present discussion of phylogenetic relationships for LT-298, L-M20, and T-M184 serves a linguistic purpose, which is to defend the *early farming dispersal hypothesis*. Here, an argument is presented that defines LT-L298 as a useful marker for deciphering the evolutionary history of the Afro-Asiatic, Indo-European and Dravidian language families. Similar arguments were previously made for haplogroups E-M96, G-M201 and J-M304 (see Chapters 5, 7 and 10).

Additionally, re-analysis of phylogenetic relationships for LT-L298, L-M20 and T-M184 seems in order because of following development: the position of the T1a-M70 mutation within the Y-chromosome phylogeny was not resolved until 2012, a decade after the publication of YCC 2002. This development revealed something unknown in 2002, that L-M20 has a sister clade, which is T-M184. This resolution of phylogenetic relationships, in turn, potentially undermines the 2002 position that L-M20 is a haplogroup. Here, the analysis requires researchers to evaluate if L-M20 and T-M184 carry the same segment of Y-chromosome diversity or, alternatively, if they independently carry a unique segment. In other words, is the "behavior" of L-M20 and T-M184 more similar to subclades like J1-M267 and J2-M172, or do they behave like haplogroups such as I-M170 and J-M304?

At this point it is important to review the concept of haplogroups that was initially presented in Section 3 of the first chapter. The standard Y-chromosome nomenclature from 2002 places Y-chromosome mutations, and more specifically, single nucleotide polymorphism, within a tree-like hierarchical structure. The theoretical Y-Chromosome Adam represents the root of a tree that eventually branches into nineteen haplogroups, such as J-M304. The haplogroups, in turn, diverge into sub-clades, e.g. J1-M267 and J2-M172. Between Adam and the haplogroups are important mutational steps (e.g. IJ-M429) that YCC 2002 identifies as "paragroups."

Focusing now on the nineteen YCC 2002 haplogroups, the label "haplogroup" denotes a single nucleotide polymorphism (or mutation) that, in turn, represents a major division within the diversity of Y-chromosome variation, or more specifically, diversity within the non-recombining region of the Y-Chromosome. YCC 2002, in their discussion of the standard nomenclature, further noted that the label "haplogroup" is "arbitrary." In practice, however, the YCC 2002 nomenclature works surprising well. For the most part, each of the YCC 2002 haplogroups represents a unique segment of Y-chromosome diversity.

Analysis of unique segments of Y-chromosome diversity considers phylogenetic relationships. Evolutionary distance between haplogroups is often, but not always, distant. For example, D-M174 and J-M304 are separated by at least eight mutational steps. Haplogroups also carry unique segments of human evolutionary history. B-M60, for example, evolved in Africa (see Chapter 3), whereas O-M175 evolved in East Asia (see Chapter 15). Another distinction between haplogroups involves expansion history, both temporal and geographical. D-M174 (see Chapter 4) records Paleolithic human expansion from Southwest Asia to East Asia during Marine Isotope Stage 3. I-M170 (Chapter 9) records human expansions into Europe during the same period. Finally, the archaeological record often supports major divisions within the diversity of Y-chromosome variation. Y-chromosome Adam (Chapter 2), for example, follows the evolution of *Homo sapiens* in Africa. J-M304 (Chapter 10) follows the expansion of the Southwest Asian Neolithic. O-M175 (Chapter 15) follows the expansion of the East Asian Neolithic.

Based on the above paragraph, haplogroups are defined by a complex interplay of phylogenetic distance, evolutionary history, expansion history, and the archaeological record. The data for L-M20 and T-M184 indicate that their evolutionary and expansion history mimic that of the J1-M267 and J2-M172. As such, they should be considered subclades of a new

LT-L298 haplogroup. Furthermore, the data suggest that although the archeological, evolutionary, and expansion history of haplogroups J-M304, G-M201 and LT-L298 are strikingly similar, they are, nevertheless, defined appropriately as haplogroups because of phylogenetic distance.

Data upon which the above position is taken partially flows from Supplementary Tables 11.1 and 11.2 which provide a survey of populations with the L-M20 and T-M184 mutations. As shown by these data, L-M20 and T-M184 surface together in several different regions: Europe, Southwest Asia, the Caucasus, and South Asia. Additionally, L-M20 is absent or virtually absent in North and Sub-Saharan Africa, whereas T-M184 has been detected in both regions. Similarly, T-M184 is absent or virtually absent in Central and East Asia, whereas L-M20 has been detected in both regions. Thus, the data suggest that L-M20 and T-M184 evolved in a single region and subsequently co-expanded into adjacent regions. A similar pattern is observed for J1-M267 and J2-M172 (see Chapter 10 and Supplementary Tables 10.2 and 10.3).

In order to further explore the co-evolution and co-expansion of L-M20 and T-M184, the reader is now directed to Supplementary Table 11.3. This table combines published data for L-M20 and T-M184 to illustrate the overall pattern of LT-L298 variation. An important observation comes from these data: the distribution pattern of LT-L298 is strikingly similar to J-M304 (see Chapter 10 and Supplementary Table 10.1). Populations with LT-L298 and J-M304 are found in the Southwest Asia, the Caucasus, North Africa, Sub-Saharan Africa, Europe, South Asia, Central Asia, and East Asia. Thus, the data point to a prehistoric population in Anatolia with variants of haplogroups E-M96, G-M201, J-M204 and LT-L298. Then during the Southwest Asian Neolithic, a catastrophic flood caused a very rapid star-like dispersal of both genes and farmers from the Black Sea.

It should be emphasized that the co-evolution of L-M20 and T-M184 in Southwest Asia, and their Neolithic expansion from this region, are supported by ancient DNA data. Lazaridis et al. (2016) reported the discovery of L1a-M27 from three ancient DNA samples taken from Areni cave in southern Armenia. These samples are about six thousand years old, and as such, date to the Neolithic. Additionally, the same study reports the discovery of a T-M184 sample from an individual that died almost ten thousand years ago at Ain Ghazal settlement in Jordan, remains that also date to the Neolithic. Finally, a T1a-M70 sample was found at a Neolithic Linearbandkeramik site in Karsdorf, Germany, which dates to about seven thousand years ago (Haak et al. 2015).

Dating estimates also support the position that L-M20 and T-M184 are signature markers of the Southwest Asian Neolithic. At this point the reader is directed to Supplementary Figure 11.1 which illustrates important phylogenetic relationships that are downstream from LT-L298. Additionally, the reader is asked to locate the L1b1-M349, T1a1-L162 and T1a2-L131 mutations. Dating results for L1b-M349 are reported by Karmin et al. (2015). According to the study, the marker evolved roughly 7.5 thousand years ago. By extension, L1a-M27 and L1a-M357 should also reflect Neolithic diversification of L-M20 variation. Turning now to T-M184, Mendez et al. (2011) provide dating results for T1a-L162 and T1a-L131, which point to the evolution of these markers roughly 11 to 14 thousand years ago. As such, they are potential Neolithic markers, like L1a-M27, L1a-M357 and L1b-M349.

Section 3. Resolving the Origins of T-M184 and L-M20.

Geneticists seem to agree that T-M184 evolved in the Middle East and expanded out of the region during the Neolithic (i.e. Mendez et al. 2011). However, the literature is much vaguer in defining where L-M20 evolved. Thus, it is necessary to further discuss the evolution of L-M20 in order to defend the position taken in Section 2 (above). Specifically, L-M20 and T-M184 co-evolved in Southwest Asia and co-expanded out of the region during the Neolithic.

Among the published reports, Lacau et al. (2012) suggest that L-M20 evolved in South Asia, and more specifically, in Pakistan. Another study, Karmin et al. (2015) report that L1-M22 evolved in South Asia and the very rare L2-L595 mutation evolved in Europe. Such a scenario seems to conform to the observed frequency and distribution of both mutations. L2-L595 has only been found in Europe (i.e. Francalacci et al. 2015). L1-M22 mutations, on the other hand, appear to have a higher frequency in South Asia relative to the frequency of the same mutations in other regions. However, it should be emphasized that even in South Asia, the frequency of L-M20 is low and hovers just around ten percent (i.e. Sengupta et al 2006; Firasat et al. 2007). As such, single region frequency data for South Asia fails to provide a convincing argument as to geographic origins of the L-M20.

In order to identify the region where L-M20 evolved, researchers should consider the overall distribution of LT-L298 variation. L-M20 and T-M184 mutations (see Supplementary Tables 11.1 and 11.2) are found in Turkey, Lebanon, and in the Caucasus, and as such, in close geographical proximity to the source population that expanded during the Southwest Asian Neolithic. Furthermore, as explained previously in Section 2 (this chapter), the distribution of LT-L298 in Eurasia and Africa roughly follows that of J-M304. Finally, as previously noted (Section 2), much of the L-M20 and T-M184 diversification occurred during the Neolithic. This is supported by ancient DNA data as well dating estimates from contemporary samples. Given these observations, the data seem more consistent with a Neolithic expansion from Southwest Asia rather than from South Asia. Otherwise, one must somehow explain a Neolithic expansion of L-M20 from South Asia to Europe, which is inconsistent with the archeological record.

Section 4. Usefulness of LT-L298 for Linguists.

In order to demonstrate the usefulness of LT-L298 as a marker for linguistic research, it is necessary to deviate slightly from the standard Y-chromosome nomenclature. Contrary to the standard nomenclature, the data suggest that LT-L298 represents a haplogroup rather than higher level paragroup mutation. Furthermore, contrary to the standard YCC 2002 nomenclature, L-M20 and T-M184 are not haplogroups. Rather, the data suggest that they are subclades within a new LT-L298 haplogroup.

At this point the reader is invited to review previous discussions of the genetic, linguistic and archaeological data for haplogroups E-M96, G-M201, and J-M304 as provide in Chapters 5, 7, and 10 respectively. Additionally, the reader is invited to review Supplementary Tables 11.4, 11.5, and 11.6 from this present chapter. Supplementary Table 11.4 sorts the L-M20 data according to language family. A similar sorting of the data is provided for T-M184 in Supplementary Table 11.5 and for LT-L298 in Supplementary Table 11.6.

The usefulness of LT-L298 as a marker for linguistic research stems from the position that this is one of four different markers that help to decipher the Neolithic expansion of agriculture from Southwest Asia, which began roughly nine thousand years ago. The Neolithic farmers of Southwest Asia must have been a population in Anatolia that had variants of haplogroups E-M96, G-M201, J-M304 and LT-L298. When these farmers expanded out of Anatolia, their genes and languages followed. The linguistic relics of this expansion include the Afro-Asiatic and Indo-European language families. As such, the *early farming dispersal hypothesis* provides a robust model of prehistoric language dispersals.

Data for T-M184 and L-M20 also help to evaluate a study published by Winters in 2010. According to the report, Dravidian languages evolved in East Africa and co-expanded with the cultivation of finger millet to India. Winters (2010) cites similarities in terminology among “Africans and Dravidians” for crops. He also supports his position by claiming that the T-M70 mutation is found in the people of East Africa and Dravidian speakers of India.

Support for the position taken by Winters (2010) potentially comes from the archaeological record. The Neolithic in South Asia has, indeed, an East African component. Furthermore, the East African Neolithic has a South Asian component. This is explained by prehistoric traders who sailed between Africa and India. As a result of this exchange, farmers in India began to cultivate finger millet and pulses such as cowpeas, crops that they had received from Africa. East Africans, in turn, received chickens as well as Asian crops such as bananas, yams and taro (see Fuller 2006; Crowther et al 2017).

According to the genetic data (see Supplementary Tables 11.4 and 11.5), both T-M184 and L-M20 have been detected in Dravidian-speaking populations. However, L-M20 appears much more frequently in these populations. L-M20, on the other hand, does not appear in Africa, whereas T-M184 occasionally surfaces in some populations of North and East Africa. However, contrary to what is asserted by Winters (2010), it seems unlikely that South Asia was the source of T-M70 variation in Africa, or that Africa was the source of the same mutation in South Asia. Again, a tremendous amount of genetic, linguistic and archaeological data, as presented here in this chapter and previously in Chapter 5, 7, and 10, all point to Southwest Asia as the source of T-M184 variation. Furthermore, these data place the likely origins of Dravidian languages in Pakistan.

Section 5. Conclusions.

LT-L298 has not received much attention because L-M20 and T-M184 generally attain a low frequency among the surveyed populations. Accordingly, the distribution of internal variation within LT-L298 is poorly understood because researchers tend to devote more time to unraveling the phylogeny of higher frequency mutations. More data would be helpful. For example, future exploration of LT-L298 variations should examine the rare L2-L595 mutation in Europe. Is this a Paleolithic or Neolithic relic? Nevertheless, based on the limited available data, L-M20 and T-M184 probably co-evolved in Southwest Asia and expanded out of the region during the Neolithic. Thus, from a big picture genetic perspective, LT-L298, G-M201, E-M96 played important supporting roles in the Southwest Asian Neolithic expansion, whereas the main actor was clearly J-M304.

Chapter 12: The KR-M526 Paragroup.

Section 1. LR-M9 versus K-M9.

KR-M526 haplogroups are important mutations for deciphering the genetic history of populations in Eurasia, Island Southeast Asia, Oceania and the Americas. At this point it is necessary to expand upon the phylogeny of LR-M9, a topic that was initially presented in Chapter 11. The reader is now invited to review Supplementary Figure 1.1. The LR-M9 mutation can be found at the lower right-hand corner on the first page of the diagram. As shown by Supplementary Figure 1.1, LT-L298 and KR-M526 are sister clades that diverge from LR-M9. LT-L298 then bifurcates into L-M20 and T-M184. From KR-M526 diverged M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207.

This resource guide designates the M9 mutation as a paragroup, an intermediate mutation between Y-chromosome Adam and the haplogroups. Hence, we designate the M9 mutation as LR-M9. The designation of M9 as a higher order paragroup represents a point of disagreement with Karafet et al. (2015). They take the position that M9 represents a haplogroup which they label K-M9. As such, they posit that K-M9 has two main subclades, K1-L298 and K2-M526. Extending their argument further, the M20 and M184 mutations become subclades of K1-L298. The P256, B254, M231, M175, M242, and M207 mutations become subclades of K2-M526. Thus, for example, the R-M207 haplogroup (from Y-Chromosome Commission 2002) becomes K2b2a2-M207 (for additional information, the reader is directed to Supplementary Figure 12.1).

The revision suggested by Karafet et al. (2015) has been partially adopted by the International Society of Genetic Genealogy (ISOGG). The organization utilizes both the Karafet et al. (2015) nomenclature and the Y-Chromosome Commission (YCC) 2002 nomenclature. However, we disagree with this revision because it erases very informative phylogenetic relationships that have very distinct evolutionary histories as well as very distinct patterns of geographic distribution. While Y-Chromosome Commission (2002) was very vague in setting standards for haplogroups, the K-M9 haplogroup proposed by Karafet et al. clearly deviates from the standard nomenclature system envisioned by the 2002 reform. In other words, the K-M9 haplogroup proposed by Karafet et al (2015) is non-standard as it encompasses far too much of the global human Y-chromosome variation. Accordingly, in conformity with the intent of the 2002 standard nomenclature, we label M9 as a higher level paragroup, which we define as LR-M9. Additionally, we retain haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207.

Section 2. The Evolutionary History of KR-M526.

As shown by Supplementary Figure 1.1, the LT-L298 and KR-M526 mutations are “sister” clades. Based on the current geographic distributions of Haplogroups LT-L298 and its downstream variants (L-M20 and T-M184), KR-M526 probably diverged from LR-M9 in the Middle East. Data from Poznik et al. (2016) suggest that this occurred roughly 45 thousand years. This dating estimate along with the contemporary distribution of haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207 suggest that KR-M526 stands as a genetic relic of an human expansions from the Levant during Marine Isotope Stage 3 (see Chapter 4: Section 1 for additional information). Other genetic relics of this expansion include haplogroups C-M130 (Chapter 6), H-M2713 (Chapter 8), and Haplogroup I-M170 (Chapter 9).

As explained previously in Chapter 11, deciphering the internal phylogeny of the LR-M9 mutation has been difficult. Taking this a step further, LT-L298, as well as haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207, all represent comparatively well-resolved LR-M9 variants. Nevertheless, among the populations of Australia, eastern Indonesia, and Papua New Guinea, a significant number of men have a KR-M526 mutation that has not been identified or sequenced. Thus they are designated as KR-M526* (Karafet et al. 2015; Nagle et al. 2016a).

Frequency data for KR-M526* is presented in Supplementary Table 12.1. Regionally, the data come from Australia, Island Southeast Asia, and Oceania. "Island Southeast Asia" encompasses the Philippines, Indonesia, East Timor, and Papua New Guinea. Oceania encompasses the eastern part of Melanesia as well as Micronesia and Polynesia. Linguistically, the data in Supplementary Table 12.1 come from populations that speak Papuan or Australian languages. The Australian family consists of 379 languages (*Ethnologue* 2018). Papuan, on the other hand, is a macro-language family of the non-Austronesian languages found in Island Southeast Asia or the Solomon Islands (i.e. Pawley 2005). As shown by Supplementary Table 12.2, the Papuan macro-family consists of over eight hundred languages that are classified into one of thirty-six language families as well as languages that are either unclassified or language isolates.

As suggested by Nagle et al. (2016a), in order to obtain better resolution of the unresolved KR-M526* data, previously sequenced samples would have to be re-sequenced for more informative markers. The fact that researchers have not better resolved KR-M526* for Australia and Island Southeast Asia may well be a question of time and money. Enormous time depth and social factors might also be a factor here, with the idea that part of the genetic trail has faded or disappeared. Turning now to the question of time depth, a recent study (Bergstrom et al. 2016) conducted sophisticated whole genome sequencing of thirteen samples that were collected from Australian aboriginals and then compared this data with that from other populations. The researchers determined, based on the structure of KR-M526*, that the unresolved mutations within this paragroup represent ancient lineages brought to Island Southeast Asia and Australia by those who settled in both regions approximately between 40 and 50 thousand years ago. This conclusion is consistent with the archeological record (see Chapter 4: Sections 1 and 2). Moreover, the conclusion agrees with data for haplogroup C-M130. As previously detailed (Chapter 6: Sections 5 and 6), the founding populations of Island Southeast Asia and Australia had the C1b2-B477 mutation. C1b2a-M38 represents *in situ* evolution of C1b-B477 in Island Southeast Asia, and C1b2b-M347 represents *in situ* evolution of C1b-B477 in Australia.

Turning now to the question of social factors that may hinder better resolution within KR-M526, Kayser et al. (2003) suggest that a long-standing tradition of warfare between the various tribes of New Guinea has reduced genetic variation among men in this area of the world. This loss of genetic diversity is akin to what known as a bottleneck effect (see Chapter 1, Section 3 for more dtails). Thus, warfare may have erased mutations that are part of KR-M526 genetic puzzle. According to the same study, another factor that may have reduced male genetic variation in Island Southeast Asia includes the prevalence of polygyny (having more than one wife).

Concerning social factors that may have reduced male genetic variation in Australia, Nagle et al (2016a) provide data taken from samples provided by 657 self-reported Australian aboriginals. More than half the men sampled by the study have a Y-chromosome haplogroup that is not indigenous to Australia. This suggests that over the last two hundred years, substantial admixture has occurred between men of European descent and Australian aboriginal women. This data suggests a potential bottleneck effect that may have reduced male genetic variation among aboriginal males. According to Nagle et al. (2016a), this may hinder better resolution of KR-M526.

Section 3. The Importance of KR-M526* for Linguists.

Previously, in Chapter 6: Section 8, C1-M347 was identified as a mutation which supports the position that language evolved at least 100 thousand years ago. Like C1-M347, KR-M526* also represents part of the indigenous genome among Australian aboriginals. Furthermore, like C1-M357, analysis of KR-M526* suggests that modern human entered Australia roughly 50 thousand years ago, and their descents remained isolated on the continent until the arrival of Europeans about two hundred years ago (Bergstrom et al 2016; Nagle et al. 2016a). Given that the initial settlement of Australian is an extension of the out-of-Africa exodus, it is very plausible that humans acquired language before leaving Africa (again, about 100 thousand years ago). The less plausible alternative scenario would posit that language evolved independently in several regions of the world.

Section 4. Conclusions.

As a matter of housekeeping we take the position that the M9 mutation is paragroup and not a haplogroup. KR-M526, a downstream variant of LR-M9, represents an important genetic relic of human expansions during Marine Isotope Stage 3. Several haplogroups eventually evolved from KR-M526. These haplogroups represent well-resolved sections of the KR-M526 phylogenetic map. Nevertheless, the literature suggests unresolved mutations within the paragroup. They are designated KR-M526*. These unresolved KR-M526 mutations are mainly found in Island Southeast Asia and Australia. For linguists this represents a salient point as better resolution of these mutations may provide greater insight into the evolution of Australian and Papuan languages.

Chapter 13: Haplogroups M-P256 and S-B254.

Section 1. Overview of Geographical and Phylogenetic Relationships.

This section features two phylogenetically close variants of paragroup KR-M526, haplogroups M-P256 and S-B254. We begin this discussion by defining our view of the geography. Island Southeast Asia consists of the Philippines, Indonesia, East Timor and Papua New Guinea. Oceania, on the other hand, consists of a broad expanse of islands in the Pacific Ocean that runs eastwards from the Solomon Islands to Rapa Nui, and southwards from the Hawaiian Islands to New Zealand. It should be noted that this description of the geography is somewhat non-standard and that terms such as Micronesia, Melanesia, and Polynesia are more standard. Additionally, the literature sometimes describes the geography as Near Oceania, Remote Oceania and Australasia. Finally, some would regard the Philippines as part of East Asia. Thus, while our geographical descriptors might be non-standard, our regional descriptions are necessary in order to facilitate an efficient delivery of the linguistic and genetic data. For additional information, see Figure 13.1 below.

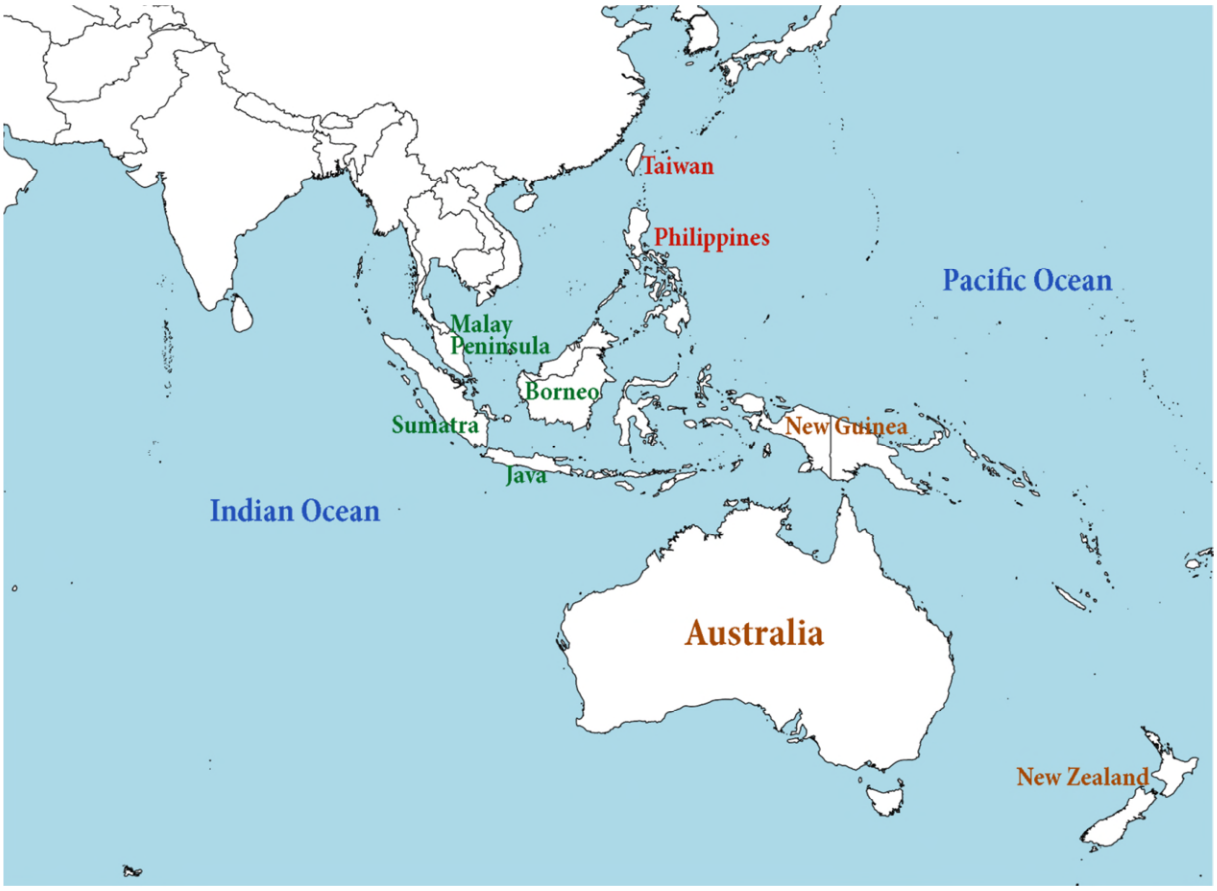


Figure 13.1. Island Southeast Asia, Australia, and Western Oceania.

The reader is now invited to review Supplementary Figure 1.1 from the first chapter. As shown on the second page of the diagram, haplogroups M-P256 and S-B254 diverge from SM-P399. The SM-P399 mutation, in turn, is a downstream variant of KR-M526. As mentioned previously in Chapter 12, haplogroups within KR-M526 stand as genetic relics of an eastward human expansion from the Levant during Marine Isotope Stage 3. This migration terminated in East Asia and Australia. Furthermore, this expansion signals the initial colonization of South Asia, East Asia, Island Southeast Asia and Australia by modern humans. Based on data from Karafet et al. (2015), SM-P399 represents initial diversification of KR-M526 in Island Southeast Asia. Then, in the same region, M-P256 and S-B254 diverged from SM-P399.

In order to further define the geographic distribution of M-P256 and S-B254, it is necessary explain the significance of the so-called "Wallace Line." This term originated within the field of botany to delineate different ecozones. Over time, it evolved into a convenient boundary that separates western Indonesia from eastern Indonesia (see, also, Supplementary Figure 6.3 from Chapter 6). The Wallace Line is especially significant for this discussion in that haplogroups M-P256 and S-B254 are found east of this boundary but are essentially absent west of the boundary (Karafet et al. 2010). Thus, the Wallace Line marks the westward limit of M-P256 and S-B254 variation in Island Southeast Asia. Oceania, on the other hand, marks the eastward limit.

For a further examination of the distribution of populations with the M-P256 and S-B254 mutations from regional and linguistic perspectives, the reader is directed to Supplementary Tables 5.13.1 and 5.13.2. The reader is also directed to Supplementary Figure 13.1 which details the internal phylogeny of M-P256, and Supplementary Figure 13.2, which details the phylogeny of S-B254. Geographically, M-P256 and its downstream variants represent significant mutations for deciphering the genetic history of Island Southeast Asia. S-B254 and its downstream variants represent significant mutations not only for the populations of Island Southeast Asia, but also a significant mutation among Australian aboriginals. Linguistically, M-P256 and S-B254 represent significant mutations among the Austronesian-speaking populations of Island Southeast Asia. Moreover, they represent significant mutations among populations that speak so-called Papuan languages. Finally, S-B254 presents data for deciphering the evolution of the Australian language family.

Section 2. Austronesian Languages.

With over 324 million speakers (Ethnologue 2018), Austronesian clearly stands as a linguistic heavyweight among the language families of the world. Moreover, this language family is distributed over a vast geographical expanse, from Madagascar to Rapa Nui. Interestingly, the linguistic and archaeological evidence point to Taiwan as the putative homeland of Austronesian languages (e.g. Bellwood 2005: 134-139; Welsch and Levine 2008; Horsburgh and McCoy 2017). This is also supported by genetic data. Since the genetic data involve variants of haplogroup O-M175, further discussion of the Austronesian homeland must wait until Chapter 15: Section 6.

Around four thousand ago, Austronesian-speaking people migrated from Taiwan to the Philippines. From the Philippines, Austronesian expanded to eastern Malaysia, Indonesia and Papua New Guinea. This expansion was carried by the so-called Lapita culture which is often identified by discarded fragments of a distinctive style of pottery. Before the arrival of the Austronesians, the people of Island Southeast Asia had spoken Papuan languages exclusively. Consequently, Papuan languages represent the indigenous linguist component of language variation in the Island Southeast Asia (e.g. Pawley et al. 2005). Moreover, they are potential linguistic relics of the initial human colonization of this region around 40 to 50 thousand years ago. Haplogroups M-P256 and S-B254, on the other hand, represent the genetic relics (see Section 1 (this paper). Thus, the widespread presence of both mutations in eastern Indonesia and Papua New Guinea among contemporary Austronesian-speaking populations is significant (see Supplementary Tables 13.1 and 13.2). Language shift and language maintenance have clearly forged language variation in Island Southeast Asia.

Section 3. Partial versus Complete Shift to Austronesian.

The Philippines, East Timor, Indonesia and Papua New Guinea were colonized by modern human at roughly the same time, 40 to 50 thousand years ago, during Marine Isotope Stage 3 (i.e. Delfin 2015). Additionally, these countries lay within the initial southward expansion zone of Austronesian languages that occurred much later. However, it should be noted that the indigenous Papuan languages of eastern Indonesia, East Timor, and Papua New Guinea managed to survive after the arrival of the Austronesians, about four thousand years ago. In the Philippines, on the other hand, Austronesian completely replaced the indigenous languages, even among the Negritos.

As previously detailed in Chapter 4: Section 6, the so-called Negrito populations found in various part of Asia, including the Philippines, are potential relic populations of the out-of-Africa expansion. The Jarawa and Onge, two Negrito populations of the Andaman Islands, retain a genetic signature of this migration. Among the Negrito populations of the Philippines, however, the data are inconclusive as to whether they still retain relic mutations. The best data for Filipino Negritos are reported by Delfin et al. (2011) and unfortunately the study utilized poor resolution markers, namely C-M130 and K-M9. The samples gathered by Delfin et al. (2011) could be re-sequenced for more informative markers. One compelling reason for taking this step is that Karafet et al. (2015) report data for the Aeta, one of the Negrito populations of the Philippines. According to the study, among the Aeta the S2-P378 mutation attains a frequency of sixty percent. This finding suggests that re-sequencing of the Delfin et al. (2011) samples may detect additional Negrito populations with S2-P378. This mutation potentially links the population of the Philippines with the rest of Island Southeast Asia, and with that provides insight about language shift and language maintenance in the region.

Focusing on the non-Negrito populations of the Philippines, it should be noted that Delfin et al (2011) provide the only source of published data for Y-chromosome mutations. The study reported data for 210 non-Negrito samples using poor resolution markers. Since 104 million people live in the Philippines (CIA World Factbook 2018), it seems that more data collection and analysis are necessary.

Section 4. Trans-New Guinea and the Early Farming Dispersal Hypothesis.

As detailed in Supplementary Table 12.2 from Chapter 12, the Papuan macro-family consists of over eight hundred languages that are classified into one of thirty-six language families. Among these language families, Trans-New Guinea occupies a unique position within the macro-language group. Trans-New Guinea is, in fact, the largest family, both in terms of number of speakers (about 3.5 million) and number of languages (almost five hundred). Additionally, the distribution of Trans-New Guinea also extends across a far greater range than the other Papuan languages. Trans New Guinea extends from the Wallace Line to Oceania, whereas the other Papuan languages are restricted to a much smaller geographic area.

A discussion of the evolution and expansion of Trans-New Guinea languages necessitates a brief discussion of Papua New Guinea in terms of its geographical location and unique topography. This country is located on the eastern half of the island of New Guinea. The low-lying coastal areas of Papua New Guinea are called lowlands, whereas the inland region is called the highlands. This contrast in altitude is the result of colliding tectonic plates which have pushed the center of the island upwards, forming a two-thousand-kilometer-long mountain chain running east to west across the islands. Here the altitude eventually climbs to four thousand meters above sea level. Extending from both sides of the mountain chain are numerous valleys (for more details, see Allen 1992).

The remote valleys of the New Guinean central highlands, along with the swamps and rainforests of the lowlands, have surely isolated human populations on the island. This pattern of isolation probably began when modern human first colonized the region, between 40 and 50 thousand years ago. Accordingly, time depth and topography provide one explanation for the extreme linguistic diversity that is found here. Topography also sets the stage for explaining the position that Trans-New Guinean languages occupy within the linguistic tapestry of Island Southeast Asia. Specifically, researchers have identified the central highlands of Papua New Guinea as the putative homeland of the Trans-New Guinea language family (i.e. Bellwood 2005: 142-145; Pawley 2005; Schapper 2017).

As previously mentioned in Paper 5.7, Hg. G, Section 1, the archaeologist Peter Bellwood formulated the *early farming dispersal hypothesis*. His hypothesis recognizes that agriculture evolved independently in several regions of the world. Taking this a step further, the hypothesis finds a good correlation between the initial expansion of early agriculture from these regions and the current distribution of many of the world's language families. Bellwood defends his hypothesis by presenting a synthesis of archaeological, botanical, climatological, linguistic and genetic data.

According to Bellwood (2005: 142-145), the *early farming dispersal hypothesis* also explains the expansion of the Trans-New Guinea language family. Denham et al. (2003) provide much of the archaeological and archaeobotanical support for this position. They collected data near a tea plantation at the Kuk Swamp in the Waghi Valley of the central highlands of Papua New Guinea, about 1500 meters above sea level. The study suggests that the agriculture transition began in Papua New Guinea about ten thousand years with the construction of drainage ditches. The intensive cultivation of taro root and bananas then evolved by around seven thousand years ago.

Both Denham et al. (2003) and Bellwood (2005: 142-145) emphasize that Holocene climate change facilitated the development of agriculture in the Papua New Guinean highlands. Higher temperatures and regular rainfall enabled people to exploit the fertile soil that accumulated in the highland valleys during the Ice Age. Their model of agricultural origins on Papua New Guinea raises an interesting question, why the Holocene populations of New Guinea preferred the highlands over the lowlands. Perhaps lowland populations were driven into the highlands by malaria where the affliction is far less prevalent.

Schapper (2017) presents a fascinating examination of early agriculture in Papua New Guinean highlands and the expansion of Trans-New Guinean languages. She disputes the traditional assumption that correlates this expansion with taro root cultivation, a staple crop of the region before the arrival of sweet potatoes. Schapper suggests that the expansion of the Trans-New Guinea language family correlates better with the cultivation of bananas and sugar cane. Her argument is partly linguistic and partly botanical. Linguistic support stems from the position that linguists are able to formulate proto-Trans-New Guinean reconstructions for banana and sugar cane. According to Schapper, reconstructions for taro root are difficult to formulate.

Botanical evidence indicates that taro root originally came from South Asia and eventually spread eastward across New Guinea during prehistoric times. This expansion pattern runs in the opposite direction as that of Trans-New Guinea expansion. Rather, the linguistic evidence points to the highlands of Papua New Guinea as the putative homeland of Trans New Guinea languages (e.g. Pawley 2005). This is based on the diversity of higher order language family sub-groups that are found in this area. Bananas and sugar cane, on the other hand are indigenous to New Guinea. As such, a westward co-expansion of Trans-New Guinean languages, bananas and sugar cane is easier to defend. Additionally, bananas and sugar cane are more versatile than taro root. Compared to taro root, bananas and sugar came grow at a greater range of altitudes and soil conditions. Their cultivation is also less labor intensive. As such, they are better suited for supporting a rapid westward population expansion that terminated at the Wallace Line.

Turning now to the Y-chromosome data, two published reports (Mona et al. 2007; Tumonggor et al. 2014) have identified the M1a1-P34 and S1a1b1-M254 mutations as the genetic signature of the Trans New Guinean expansion. Particularly persuasive are the dating estimates for the M1a-P34 and S1a-M245 as provided by Mona et al. (2007). The study reports that both mutations evolved roughly seven thousand years ago in Papua New Guinea. Additional support is found in Supplementary Table 13.3 (Papuan and S1a-M254) and Supplementary Table 13.4 (Papuan and M1a-P34). Based on the tables, the distribution of S1a-M254 and M1a-P34 mutations conforms to a model of Trans-New Guinea language family origins in the highlands of Papua New Guinea and westward expansion of these languages to the Wallace Line, and an eastward expansion to Oceania.

Close examination of the data from the supplementary tables referenced in the preceding paragraph has revealed, nevertheless, a serious deficiency in the quality of available data. Almost all the data come from low-lying coastal area. The only Y-chromosome data reported for the highlands of Papua New Guinea come from 31 samples sequenced by Kayser et al. (2006). Unfortunately, very little is known about the individuals who furnished the samples. These samples were initially collected from placental tissue by Stoneking et al. (1990) for an early mtDNA study. According to the 1990 study the tissue samples came from several villages in the highlands from people who spoke "non-Austronesian." Their "non-Austronesian" language or languages probably belong to the Trans-New Guinea language family. This is based on a comparison of language maps prepared for Papua New Guinea by *Ethnologue* (2018) with a map furnished by the 1990 study.

Another disturbing matter that needs to be brought to the reader's attention concerns the M1a-P34 mutation. On January 9, 2017, the International Society of Genetic Genealogy (ISOGG) removed M1a-P34 from the Y-DNA haplogroup tree because the mutation fails to meet their "quality guidelines." This development is problematic because seven different studies report P34 data for 2,496 men and now researchers lack certainty as to where the mutation is positioned within the phylogeny of haplogroup M-P256.

Finally, it should be emphasized that lack of research for Papuan and Trans New Guinean languages stands as a serious deficiency within the field of linguists. Andrew Pawley writes the following:

> There is not a single linguist whose primary research field is Papuan historical linguistics. Only a handful of linguists are active in [Trans-New-Guinean] historical studies. It might be said that studies of the Trans New Guinea family are about where Indo-European studies were in the 1820s, in the days of Rask and Grimm, but with the prospect of having only a tiny fraction of the manpower that was available for the study of Indo-European (2005: 99-100).

Nevertheless, despite all these limitations, the available genetic, archaeological, botanical, and linguistic data still point to the *early farming dispersal hypothesis* as a robust model for explaining the prehistoric expansion of Trans-New Guinean languages.

Section 5. Other Informative Markers for Papuan Languages.

At this point we transition away from the S1a-B254 and M1a-P34 mutations and present other downstream variants of S-B254 and M-P256. Our goal is to provide additional informative markers for deciphering the prehistory of Papuan languages. Supplementary Table 13.5 presents data for S1a2-P79. This mutation represents a potentially informative marker for the Papuan languages of the Bismarck Archipelago of Papua New Guinea, including the East New Britain and Yele-West New Britain language families. Additionally, S1a-P79 data are reported for two language isolates, Sulka and Kuot.

Supplementary Table 13.6 reports data for M1a2a-P87. This marker has been found in populations that speak languages classified within the East New Britain¸ Yele-West New Britain, Central Solomons, and North Bougainville language families, as well as the isolates Sulka and Kuot. As shown by Supplementary Table 13.7, the M2-M353 mutation is found among speakers of language belonging to the Central Solomons family. Finally, Supplementary Table 13.8 reports data for the M3-P117 mutation, which might be an informative marker for the Central Solomons, East New Britain¸ and Yele-West New Britain language families, and the Sulka language isolate.

Section 6. Australian Languages.

Researching the genetic history of Australian aboriginals has been problematic. Holst Pellekaan (2013) provides an explanation stating that aboriginal Australians are reluctant to participate in genetic studies due to historical mistrust between themselves and Europeans. Interestingly, a similar problem has arisen with respect to genetic studies that focus on Native Americans (see Chapter 16: Section 2). Of course, the paucity of data for aboriginal Australians is extremely unfortunate for linguists because we lack genetic data for over three hundred languages classified within the Australian language family. Furthermore, according to Ethnologue (2018) only 185 Australian languages are still spoken and many of these living languages face an uncertain future. Finally, as noted earlier in Chapter 6: Section 8, the genetic history of Australian aboriginals helps to define when language evolved as a human behavior.

Most of the Y-chromosome data for Australian aboriginals come from Nagle et al. (2016a). Their study certainly represents a step in the right direction. However, the quality of their data is mediocre, a point that the study seems to concede. Good quality data focuses on populations and provides ethno-linguistic details. Nagle et al. (2016a), however, gathered their data from four different databases from men who identify themselves as aboriginal Australians. Unfortunately, the regional and language affiliations of the men are missing.

As noted earlier in Chapter 6: Section 6, C-M347 has emerged as a unique Australian-specific mutation that represents a genetic artifact of the initial human settlement of Australia 40 to 50 thousand years ago. Undefined mutations within KR-526 are also a relic of this migration (see Chapter 12: Sections 2 and 3). In their 2016a study, Nagle and others identify the S1a1a1-P308 mutation as another genetic artifact of this migration. According to the study, this mutation attains a frequency of about twelve percent among the Australian aboriginals (see, also, Supplementary Table 13.9).

Nagle et al (2016a) assert, based on their analysis of the Y-chromosome data, that after humans colonized Australia between 40 and 50 thousand years ago, the ancestors of contemporary aboriginals remained isolated from the rest of the world until the arrival of Europeans in the eighteenth century. It should be noted that data from mitochondrial DNA (mtDNA) also support this conclusion (Nagle et al. 2016b). The significance of the mtDNA data stems from the fact that it is inherited maternally, and Y-chromosome DNA is paternally inherited. Thus, researchers can eliminate the possibility of Pleistocene male origins among the Australian aboriginals and subsequent female mediated gene flow at a much later time. Taking this a step further, both mtDNA data and Y-chromosome data support the following argument: Australian languages have roots that potentially extend deep into the prehistory of modern humans.

During the 40 to 50-thousand-year period that Australian populations were isolated from the rest of the world, very limited gene flow may have occurred via the Torres Strait, where the distance between Papua New Guinea and Australia narrows to about 150 kilometers. This is based on presence of the M1-M4 mutation that is detected in about one percent of the Australian aboriginals, as reported by Nagle et al (2016a). According to the study, the significance of M1-M4 among Australian aboriginals needs further analysis. A future investigation of this matter will require the collection of more data from Queensland, the Australian state closest to Papua New Guinea.

Section 7. Conclusions.

Haplogroups M-P256 and S-B254 have emerged as important markers for understanding the genetic history of populations in Island Southeast Asia and Australia. For linguists, these mutations are important markers for deciphering the prehistory of Papuan, Australian and Austronesian languages. Data from both mutations suggest that language shift and language maintenance have clearly forged language variation in Island Southeast Asia. The indigenous Papuan languages of eastern Indonesia, East Timor, and Papua New Guinea managed to survive after the arrival of the Austronesians. However, Austronesian completely replaced the indigenous languages of the Philippines.

The M1a-P54 and S1a-M254 mutation are informative markers for explaining the expansion of Trans New Guinea languages. These genetic data, along with the archaeological, botanical and linguistic evidence, suggest that this expansion conforms to the *early farming dispersal hypothesis.*

Finally, the S1a-P308 mutation represents an important marker for deciphering the genetic history of aboriginal Australians. This, in turn, provides a convenient scale to assess the minimal age of human language. We witness the potential evolution of over 300 indigenous Australian languages from a language spoken by a very ancient population. This ancient population descended directly from the initial out-of-Africa exodus. Thus, when our ancestors first left Africa about 100 thousand years ago, among the tools they carried for the journey was language.

Chapter 14: Haplogroup N-M231.

Section 1. Overview.

The N-M231 haplogroup and its downstream variants help to decipher genetic diversity throughout a vast area that consists of Northern Eurasia, the Baltic Region, Scandinavia and Eastern Europe. Additionally, the same data help to resolve the prehistory of several different language families. In Section 2, the discussion of haplogroup N-M231 begins with an analysis of phylogenetic relationships. In the section that follows (3) the reader encounters the *reindeer hypothesis*. The hypothesis associates a rapid bi-directional expansion of N1a-M46 about five thousand years ago with better reproductive success among reindeer herders. Finally, as detailed in Section 4, better reproductive success among reindeer herders fueled an expansion of the Uralic language family. Languages such as Komi, Finish and Estonian stand as Uralic relics of this expansion. This expansion also involved Uralic-speaking populations that shifted to a new language. Thus while N1a-M46 represents a genetic signature of the Uralic family, the same mutation helps to decipher the prehistory of Indo-European, Tungusic, Turkic, Mongolic, Chukotko-Kamchatkan, and Eskimo-Aleut.

Section 2. Phylogenetic Relationships.

2.1. Origins of Paragroup NO-M214.

At this point the reader should review Supplementary Figure 1.1 from the first chapter. As shown on the second page of the figure, the NO-M214 mutation is a downstream variant of KR-M526. Karmin et al. (2015) and Ilumae et al. (2016) both suggest that NO-M214 arose roughly 42 thousand years ago. To determine where this occurred, it is necessary to revisit the out-of-Africa model that was previously introduced in Chapter 4. The model suggests that *Homo sapiens* migrated from eastern Africa into the Levant about 130 to 100 thousand years ago during Marine Isotope Stage (MIS) 5, around the beginning of the last Ice Age. For a period of between 50 and 80 thousand years, human populations in the Levant expanded. During Marine Isotope Stage 3, about 45 to 50 thousand years ago, a temporary amelioration of the widely fluctuating Ice Age climatic conditions facilitated human migration out of the Levant. This migration signals the arrival of modern human in East Asia as well as Europe.

Recent interpretations of the archaeological, genetic and paleo-climatological data report that *Homo sapiens* colonized East Asia via a single migration route during Marine Isotope Stage 3. According to these reports, this so-called “southern route” initially followed the coastline of southern Asia. In southeastern Asia, some proceeded northwards along the coastline of East Asia. Other traveled south into Island Southeast Asia and Australia (see Mellars 2006; Pope and Terrell 2008; Stoneking and Delfin 2010; Oppenheimer 2012; Karafet et al. 2015). From a Y-chromosome perspective, genetic relics of the southern migration model include downstream variants of haplogroups D-M174, C-M130, M-P256 and S-B254. Since NO-M214 diverged from KR-M526, this suggests that NO-M214 also evolved somewhere in southeastern Asia, like M-P256 and S-B254. Such a scenario is posited by Karafet et al. (2015). However, the position taken by these researchers is inconsistent with the ancient DNA data for C-M130 and NO-M214. Specifically, the single southern migration model fails to explain why C-M130 and NO-M214 were part of the Central Asian and European gene pools during the Paleolithic.

At this point the reader is invited to examine Supplementary Table 14.1 which provides a survey of Paleolithic Y-DNA samples. As shown by the table (see Reference Sample No. 1), the Ust'-Ishim man died about 45 thousand years in western Siberia in what is now the Omst Oblast of Russia. Researchers determined that he had the NO-M214 mutation. Similarly, the NO-M214 mutation was sequenced from Paleolithic remains found at the Peştera cu Oase cave in Romania (Reference Sample No. 2). Supplementary Table 14.1 also indicates that C1-F3393 mutations were found in Paleolithic remains from Eastern, Central and Western Europe (See Reference Samples 3, 4 and 5). These data, along with the paleo-climatological record, suggest that modern humans were drawn to Central Eurasia about 45 to 50 thousand years ago because of the availability of high-quality food. According to Ricankova et al. (2014), during the last Ice Age this area consisted of a tundra-steppe ecosystem, the so-called "mammoth steppes." As implied by the term, this ecosystem supported a variety of large herbivores that not only included mammoths, but also reindeer, woolly rhinoceroses, wild horses, bison, and wild goats (Dolukhanov 2003; Gordon 2003).

The genetic and climatological data, as detailed above, point to a population living somewhere near the Black Sea and Caspian Seas during Marine Isotope Stage 3. The population split. One group may well have followed the Euphrates and Tigris Rivers to the South Asian coast and points beyond. Motivated by the prospect of harvesting large herbivores, the other group migrated onto the mammoth steppes of Central Eurasia, perhaps through the Caucasus, or perhaps somewhere east of the Caspian Sea. This expansion into Central Eurasia potentially consisted of men with the KR-M526, C1-F3393 and C2-M217 mutations. Somewhere on the steppes, NO-M214 evolved from KR-M526. Then another population split occurred. One group carried NO-M214 and C1-F3393 from Central Eurasia westwards to Eastern and Western Europe. The other carried NO-M214 and C2-M217 eastwards from Central Eurasia to China and Mongolia.

2.2. The Origins and Initial Diversification of N-M231.

It should be noted that despite phylogenetic closeness, the sister clades N-M231 and O-M175 have very different contemporary geographic distributions. Haplogroup O-M175 represents an important marker in East Asia, South Asia and Oceania, whereas haplogroup N-M231 attains a frequency of about only about six percent in East Asia (Zhong et al. 2011). See, also, Chapter 15: Section 1. Additionally, N-M231 is almost absent in South Asia and Oceania. Instead, haplogroup N-M231 stands as a marker of North Eurasian populations with a geographic distribution that extends over a vast territory from the Pacific Ocean in the East to the Atlantic Ocean in the West. O-M175, on the other hand, is virtually absent in Northern Eurasia.

Haplogroup N-M231 and its sister clade, O-M175, evolved from NO-M214 about 38 thousand years ago (Poznik et al. 2016: Supplementary Table 10). The genetic and paleoclimatological evidence (e.g. Shi et al. 2013) suggest that this occurred in China. Additional support comes from a 2015 report published by Hu et al. Haplogroup N-M231 has two main internal divisions, N1a-F1206 and N1b-F2930. According to Hu et al. (2015), both mutations evolved roughly 16 thousand years ago in China. N1b-F2930 eventually became the dominate haplogroup N-M231 variant in East Asia (see Supplementary Table 14.1). N1a-F1206, on the other hand, represents N-M231 variation in Northern Eurasia (see Supplementary Table 14.3).

From a climatological perspective, the evolution of N1a-F1206 and N1b-F2930 occurred close to the end of the last Ice Age and the beginning of the Holocene. N1b-F2930 remained in East Asia. N1a-F1206 eventually migrated northwards. This diversification of N-M231 into North Eurasian and East Asian variants conforms to an expansion model presented by Xue et al. in their 2006 study. Their report analyzes haplogroup and short tandem repeat data from the Y-chromosome. The researchers found a significant population expansion around the time of the Last Glacial Maximum in northern China. However, populations in central China started to expand much later, around the beginning of the Holocene. According to the researchers, Pleistocene populations in northern China expanded because they were able to exploit the abundant food resources on the so-called "mammoth steppes" until the beginning of the Holocene.

Section 3. The Reindeer Hypothesis and N1a-M46 Variation in Northern Eurasia.

3.1 Overview of the Genetic Data.

Turning now to the genetic data, the reader is directed to Supplementary Figure 14.1, which provides a phylogenetic overview of haplogroup N-M231. The format used for this figure departs slightly from that used for phylogenetic charts presented in previous sections. This was necessary because the internal structure of N-M231 is unusually wide and deep. To fit the phylogeny of N-M231 on a single page, the usual statement that typically accompanies the cladistic and mutational identifiers was omitted. Instead, the chart lists several different levels which correspond to the number of characters in the cladistic identifier. For example, N1a1a2 stands at Level 6 because it has six characters. These levels, as the reader shall soon see, carry a discussion of the data.

The N1a-F1206 mutation is found at Level 3 in Supplementary Figure 14.1. The N1a-P43 mutation, a downstream variant of N1a-F1206, is found at Level 5. Data for this mutation appears frequently in published reports. However, N1a-P43 is found in populations over a vast geographical expanse, from the Baltic region to eastern Siberia (see Supplementary Table 14.4). As such, the mutation is not particularly informative. Better resolution of its downstream variants is needed.

Found at Level 3 is the N1a-M46 mutation. Researchers (Zerjal et al.) reported discovery of the mutation in 1997 and since then its frequency among Eurasian populations has been published in numerous reports (see Supplementary Table 14.5). In these reports, N1a-M46 is often called the "Tat" mutation. The term "Tat" describes an unusual category of genetic mutations that are phylogenetically equivalent to M46. Among the most significant N1a-M46 study is that published by Ilumae et al. (2016). The researchers report that N1a-M46 evolved roughly 13 thousand years ago. This dating estimate, along with its contemporary geographic distribution, suggests that N1a-M46 is a genetic relic of the human colonization of Northern Eurasia during the Holocene. The same study also reports several informative downstream markers for mutation that clarify the evolutionally history of the mutation.

At this point the reader may want to review Supplementary Figure 14.1 and the phylogenetic overview of haplogroup N-M231. Within the figure, informative N1a-M46 variants are found at Level 8 (N1a-B211), Level 10 (N1a-Z1936 and N1a-M2019) and Level 11 (N1a-VL29, N1a-B479, N1a-F4205, and N1a-B202). The N1a-B211 mutation is found in

Northern Eurasia and Eastern Europe among populations that speak Uralic or Turkic languages (see Supplementary Table 14.6). As shown by Supplementary Table 14.7, N1a-Z1936 is found in Scandinavia among Finns and Sami. It is also found among Russians, Veps and Karelians in the Baltic region, and in Northern Eurasia among Turkic-speaking Tatars and Uralic-speaking Komi. N1a-M2019 is found in Northern Eurasia among Tungusic and Turkic speaking populations (see Supplementary Table 14.8). Among Estonians, Latvians and Lithuanians, N1a-VL29 (see Supplementary Table 14.9) attains a significant frequency. The same mutation is found among the Uralic-speaking populations of Scandinavia, Eastern Europe, and Northern Eurasia. N1a-B479 appears to be the genetic signature of the Nanai, a Tungusic-speaking people in Eastern Siberia (see Supplementary Table 14.10). The N1a-M4205 mutation (see Supplementary Table 14.11) appears to be a unique mutation of Mongolic-speaking populations in East Asia and Northern Eurasia. Finally, as shown by Supplementary Table 14.12, N1a-B202 is found in Eastern Siberia among the Chukchi, Koryaks, and Yupik.

The informative N1a-M46 variants, as just described, exhibit the following characteristics: (1) they all evolved between four and five thousand years ago; (2) their contemporary geographic resulted from a bidirectional east-to-west or west-to-east expansion across Northern Eurasia of genetic mutations that are very close phylogenetically; (3) the expansion terminated at geographic points that are distant, approximately 5,000 miles apart; and (4), their expansion across Eurasia was rapid. Ilumae et al. (2016) suggest this expansion was driven by the Seima-Turbino cultural phenomenon and metallurgy. However, an alternative *reindeer hypothesis* is better supported by the genetic, anthropological and paleo-climatological evidence. A convergence of these data sources suggests that the expansion was driven by a cultural adaptation which vastly improved human reproductive success. This cultural adaptation is the domestication of reindeer.

3.2. The Significance of Reindeer Domestication.

To underscore the historical significance of reindeer as a source of food source for *Homo sapiens*, Gordon (2003: 15) writes the following:

> It [reindeer] dominated numerically and geographically, and was used by people more intensively than any other animal. It was more important than North American or Ice-Age European wild cattle, bison, mammoth, mastodon or horse. It was more important than seals and whales in all the oceans; more important than red deer, black and white-tailed deer, moose and elk. It was more important than the great African herds of antelope, zebra and gazelle. It and its hunters occupied half the land north of the equator.

The behavior of wild reindeer involves migration across the tundra especially during the spring and summer. Several factors influence this behavior including the presence or absence of forage, the depth of snow, mosquitos, and the freezing and thawing of rivers (Baskin 1986). Prehistoric hunter-gatherers intercepted herds of migrating reindeer at strategic locations, such as river crossing (e.g. Baskin 2003). While reindeer were an excellent source of food for hunter-gatherers, the unpredictable migration cycle of this animal meant that they were not always a dependable source of food. The domestication of this animal obviously changed this situation, and reindeer meat became a reliable year-round source of nutrition.

Gordon in his 2003 paper addresses the correlation between fertility among hunter-gatherers and the migration cycle of reindeer. According to Gordon, women require minimum of twelve percent body fat to conceive, and eighteen percent body fat to carry a child until birth. Based on data he collected among hunter-gatherers that harvest reindeer, women within these groups generally conceive in the late spring and their children are born at the end of the following winter. This pattern of conception and birth among female hunter-gatherers correlates with the migration of reindeer from the forests in the early spring. They begin a trek northward onto the tundra, where calves are born and where they feed from the lichen that grows abundantly in this area. On the open tundra hunter-gatherers can easily harvest this animal, which in turn provides the nutrition need by women for conception. When reindeer return to the forest in the fall, food becomes scarcer for human societies that hunt these animals, which hinders conception among women.

The previous discussion of human female fertility and the migration cycle of reindeer are potentially significant in that it may explain why population expansions occurred in the last 5 thousand years among several different cultures that domesticate reindeer. These expansions correlate well with the domestication of reindeer because domestication broke the previous cycle of feast and famine among hunter-gatherers. A more dependable year-round source of fat and protein emerged, which meant better nutrition for the group. This, in turn, resulted in greater fertility for women and a corresponding reduction in childhood mortality. Better reproductive success, and the need to move domesticated herds of reindeer to “greener pastures” to feed a rapidly expanding human populations, may have facilitated the rapid bi-directional expansion of haplogroup N as suggested by the data provided by Ilumae et al. (2016).

3.3. Where Domestication first Occurred.

Two different models of reindeer domestication have circulated among experts in this area (e.g. Gordon 2003). The diffusion theory suggests that reindeer were initially domesticated in single location and this practice later spread through cultural contact. The evolution theory suggests domestication arose independently in several different areas. Two reports, Mirov (1945) and Gordon (2003), favor a single location near the source of the Yenisei River in the Tuva Republic of Russia. Both reports are based on anthropological data including prehistoric reindeer pictographs found in this area.

Support for the diffusion theory partially stems from genetic data gathered from reindeer. Røed et al. (2008) analyzed mitochondrial and micro-satellite markers gathered from wild and domesticated herds across the Eurasian landmass. According to their data, European and Asian reindeer have a common Paleolithic origin that diversified genetically during the Holocene. Furthermore, the data reflect independent domestication of reindeer in Scandinavia and in northern Russia. However, the researchers could not determine whether domestication in northern Russia occurred in a single location of this vast region or, alternatively, if domestication arose independently in several different locations. According to their report, much of the genetic history of reindeer in northern Russia has been erased by a long-standing practice of augmenting domesticated herds with reindeer taken from wild herds. Nevertheless, despite the limitations of their work, these researchers still offer an important clue as to where reindeer were first domesticated. They found wild reindeer from the tundra were the source of domesticated reindeer. This is an important conclusion as it eliminates a sub-species of reindeer that evolved during the Holocene in the boreal forests (or taiga) south of the Eurasian tundra. Rather, reindeer domestication occurred further north near the Arctic Circle.

The anthropological perspectives provided by Mirov (1945) and Gordon (2003), along with the reindeer genetic data provided by Røed et al. (2008), suggest that reindeer were initially domesticated on the Taymyr Peninsula of northern Russia. Support for this position is that one of the largest contemporary herds of wild reindeer in the world is found on the Taymyr Peninsula (e.g. Pavlov 1994; Kolpaschikov 2015). See, also, Figure 14.1 below. Near this location in northern Russia, the Yenisei empties into the Arctic Ocean. The wild reindeer of the Taymyr Peninsula migrate northwards onto the peninsula in the spring. When autumn approaches, they migrate southwards to spend the winter along the Yenisei River in the boreal forest (Baskin 1986).

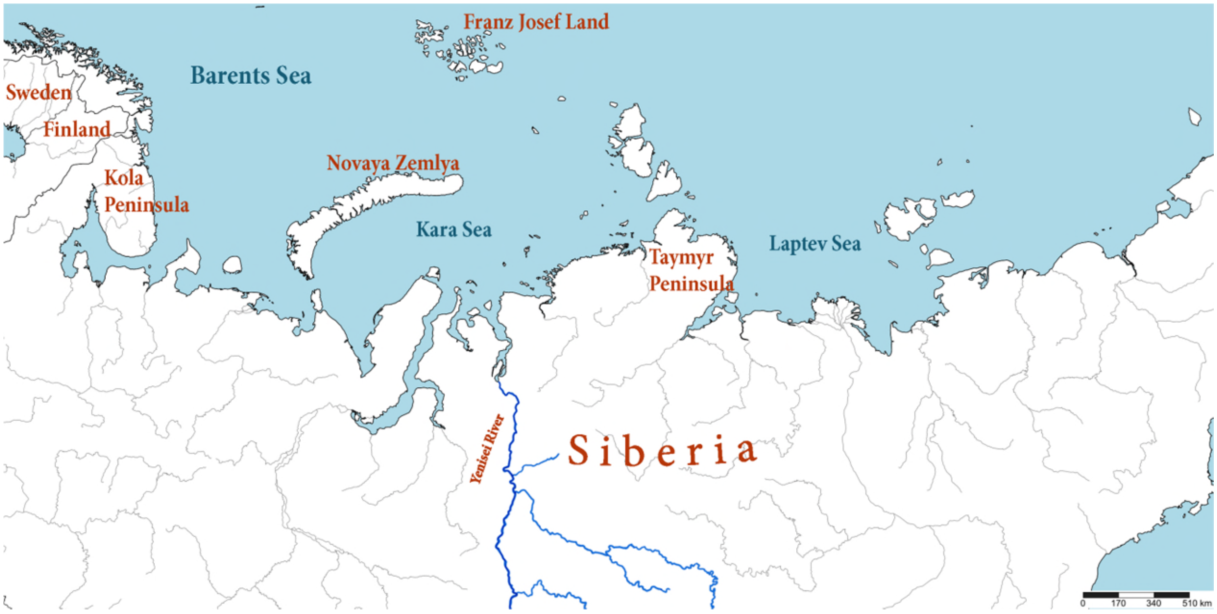


Figure 14.1. Taymry Peninsula and Yenisei River.

The idea that reindeer were initially domesticated on or near the Taymyr Peninsula follows and expands upon the paleo-climatological and genetic data that was previously detailed. During the Last Ice Age the so-called “mammoth steppes” of Eurasia provided an ideal habitat for large herbivores. Human populations throughout the Northern Hemisphere exploited this resource (see, also, Chapter 16: Section 3). Suddenly, with the onset of the Holocene and warmer weather about 14 thousand years ago, the mammoth steppes began to retreat northwards from the fortieth parallel (Ricankova et al. 2014). When the tundra began to retreat with the onset of the Holocene, the reindeer of the former Ice Age mammoth steppes retreated with the tundra. Hunter-gatherers in northern China followed the reindeer. The genetic data suggest that these hunter-gatherers had the N1a-M46 mutation. According to Ilumae et al. (2016), the N1a-M46 mutation evolved at the onset of the Holocene, around 13 thousand years ago. Hu et al. (2016) place the origins of the N1a-M46 mutation in northern China or Mongolia. Since the source of Yenisei River is found close to this location, it appears the expansion followed the receding tundra line and reindeer along this important waterway of Siberia.

3.4. When Domestication First Occurred.

As previously suggested, around 14 thousand years ago hunter-gatherers began to migrate northwards from China or Mongolia. These hunter-gatherers carried the N1a-M46 mutation towards the Taymyr Peninsula. It appears based on paleoclimatological data that this northwards migration of genes and reindeer hunters to the Taymyr Peninsula ended around five thousand years ago. At this point the tundra had finally receded to the Arctic Circle (Binney et al. 2016). The termination of this migration, as suggested by the climatological data, correlates well with the beginning of reindeer domestication as suggested by an anthropological perspective of the data (Gordon 2003). Human genetic support for this position stems from data provided by Ilumae et al. (2016). Informative N1a-M46 markers reported by the study (i.e. N1a-B211, N1a-Z1936, N1a-M2019, N1a-VL29, N1a-B479, N1a-F4205, and N1a-B202) all evolved within the last five thousand years. Taking this a step further, these mutations are genetic relics of improved reproductive success that fueled human expansions.

Turning now to human genetic data, the reader is invited to review Supplementary Table 14.13. The table provides the frequency of N1a-M46 among populations for which reindeer domestication is a current or recent practice. Particularly striking about the table is that it represents an excellent cross section of populations along the entire northern Eurasian landmass, from Scandinavia to the Bering Sea: Sami (Scandinavia); Komi (Volga Uralic Region), Selkups, Nganasans, Nenets, Khanty and Dolgans (Western Siberia); Sojots, Yakuts, Evenki, Even, Yukaghir and Dolgans (Central Siberia); Tuvans and Tofalars (Southern Siberia); Chukchi and Koryaks (Eastern Siberia). Focusing now on reindeer herders and the informative N1a-M46 mutations that evolved in the last five thousand years, Supplementary Table 14.14 indicates that N1a-B211 attains a low to moderate frequency among the Komi. N1a-Z1936 attains a low frequency among Komi, Nenets, Dolgans, and Nganasans (Supplementary Table 14.15). Among the Sami, the same mutation attains a moderate frequency. Turning now to Supplementary Table 14.16, the N1a-M2019 mutation attains a heavy frequency among Yakuts, Dolgans, Even and Evenks. Supplementary Table 14.17 indicates that the N1a-VL29 mutation attains a moderate frequency among Sami, Nenets and Komi. Finally, as shown by Supplementary Table 14.18, the N1a-B202 mutation attains a moderate frequency among Koryaks, and a heavy frequency among the Chukchi. As previously suggested, this dispersal pattern of N-M231 mutations is consistent with a demographic model that posits rapid human expansions and better reproductive success.

3.5. Summary of the Reindeer Hypothesis.

As noted above, downstream variants of the N1a-M46 (or N1a-Tat) mutation provide especially informative markers for deciphering the spread of haplogroup N-M231 across Northern Eurasia, as well as the Baltic and Scandinavia. The *reindeer hypothesis* explains the expansion, which as noted earlier, began about five thousand years ago. The hypothesis follows warmer climate with the onset of the Holocene. The tundra receded northwards. The wild herds of reindeer followed the tundra. Hunter-gatherers followed the reindeer. Then around five thousand year ago the northward expansion of tundra arrived at its presence location at the Arctic Circle and the reindeer hunter-gatherers encountered a geographic dead-end. At this point an important cultural adaptation occurred, the domestication of reindeer. This animal now became a reliable year-long source of food instead of a seasonal source of sustenance. Population pressure among the herders, as well the need to provide forage for the domesticated herds of reindeer, resulted in human migrations both eastwards and westwards. The eastward expansion eventually encountered a geographic cul-de-sac at the Bering Sea and similarly the westward expansion terminated at the North Sea.

Section 4. Significance of N-M231 for Linguists.

4.1. Uralic and the *Early Farming Dispersal Hypothesis*.

Ethnologue (2018) reports thirty-seven Uralic languages. According to the same source, around 20 million people speak a Uralic language. Roughly half this figure belongs to Hungarian. Finnish, with 5.2 million speakers, and Estonian, with one million speakers, also represent Uralic "heavyweights." The moderate to heavy frequency of downstream N1a-M46 mutations among many Uralic-speaking populations is striking. These data provide additional support for the *early farming dispersal hypothesis*, the idea that the Neolithic transformation offers a good correlation between the initial expansion of early agriculture and the current distribution of many of the world's language families. As the reader may recall, the Neolithic transformation involved farmers who cultivated crops, and pastoralists who domesticated animals. A particularly strong example of early pastoralism and language expansion occurred in Southwest Asia and North Africa. Here the domestication of sheep and goats fueled an expansion of Afro-Asiatic languages (see, Chapter: 10: Sections 2 and 3). Similarly, in northern Eurasia the domestication of reindeer fueled an initial expansion of Uralic languages. The expansion of Afro-Asiatic and Uralic languages also produced a genetic "scar." In Southwest Asia and North Africa, the J1-M267 mutation represents a genetic relic of the expansion of Afro-Asiatic (see Chapter 10). For Uralic, the genetic artifact is N1a-M46.

4.2. Samoyed and Finno-Ugric.

Genetic, archaeological, anthropological and linguistic evidence suggest that Nenets are potentially the closest contemporary representatives of a prehistoric population that brought Uralic languages to Scandinavia. According the anthropological perspective (Mirov 1945) Samoyed-speaking populations were the first cultural group to domesticate reindeer. This is linguistically significant as the Nenets language is classified within the Samoyed branch of Uralic languages. Turning back to the archaeological perspective, the Nenets live near Taymyr Peninsula where reindeer were initially domesticated (see Section 3, this paper). Additionally, the Nenets have traditionally herded reindeer.

Turning now to the genetic evidence, N1a-VL29 and N1a-Z1936 are among the informative mutations reported by Ilumae et al. in 2016. Both mutations are present among the Nenets of northern Siberia. The same mutations are found among the Finns and Sami of Scandinavia (see Supplementary Tables 14.7 and 14.9. The arrival of Uralic languages in Scandinavia is associated with the Comb Ceramic Culture. The archeological record suggests that this occurred around five thousand years ago (e.g. Siiräinen 2003). This interpretation from the archeological record is significant as it provides additional evidence that associates the expansion of Uralic languages with reindeer domestication in Northern Eurasia (Section 3). In other words, the dating estimates are very close. Turning again to the linguistic data, language classification positions the Samoyedic and Finno-Ugric branches as the two main divisions within the Uralic language family (e.g. Austerlitz 2009).

The above synthesis of genetic, linguistic, archaeological, anthropological and perspectives suggest that a proto-Uralic-speaking population lived somewhere close to the Taymyr Peninsula. The Samoyedic branch represents linguistic diversification among those that stayed. Finno-Ugric represents diversification among those that left.

4.3. Hungarian.

The N1a-M46 mutation maintains an astonishingly high frequency among many populations speaking a Uralic language, such as ninety-three percent of Nenets, fifty-four percent of Finns, and thirty-one percent of Estonians (e.g. Ilumae et al. 2016). However, despite speaking a Uralic language, haplogroup N-M231 and its downstream variants are almost completely absent among Hungarians (e.g., Völgyi et al. 2008). While the genetic history of Finns and Saami point to population expansion as having the potential to alter the linguistic landscape of a region, the genetic history of Hungarians reflects that population expansion is not a prerequisite for language expansion. Such a scenario for the Hungarians language agrees with the historical record. A relatively small population from Central Asia, the Magyars, invaded Europe in the fourth century. Later a much larger Central-European population shifted to the language of their conquerors from western Siberia.

Although Magyars contributed very little to the contemporary Hungarian gene pool, a recent genetic study has identified the Mansi people of western Siberia as a potential source population for the Magyar invasion. According to Fehér et al. (2015), the N1a-L1034 mutation links contemporary Mansi and Hungarians. From a linguistic perspective, Austerlitz (2009) views the Mansi and Hungarian languages as essentially two different sub-branches of Ugric, and as such, they are linguistically close. Thus, the linguistic and genetic data may agree.

4.4. North Germanic.

Within Scandinavia the traditional pattern of language variation consists of languages classified as North Germanic, Finnic and Sami. N1a-M46 appears to be the genetic signature of Uralic-speaking Finns and Sami based on the moderate to heavy presence of the mutation in both populations and its comparatively low frequency among ethnic groups that speak a North Germanic language. For example, N1a-M46 is virtually absent among Danes (Sanchez et al. 2004). Among the Norwegians, less than three percent have the mutation (Dupuy et al. 2006). Among Swedes, the figure stands between ten and fourteen (Karlsson et al. 2006; Lappalainen et al. 2006). The I1a-M253 mutation, on the other hand, seems more evenly distributed among all the Scandinavian populations. It is present in about one third of Finns and Saami (Tambets et al. 2004; Lappalainen et al. 2006). Similar frequencies are detected among Danes, Norwegians and Swedes (see Chapter 9: Section 7).

The data for N1a-M46 and I-M253 in Scandinavian populations reflects that language contact is very much a part of linguistic evolution in this area of the world. Indeed, Finnish borrowings from Germanic supplement the archaeological and genetic data to suggest admixture between Uralic and Germanic tribes in prehistoric Scandinavia. For example, Fromm (1977) argues that these loanwords may point to the presence of the Germanic tribes in central Sweden during the Bronze Age, roughly 3,000 years ago. Additionally, since Finnish has changed relatively little over the two past millennia, the Germanic borrowings in Finnish are thought to provide a well-preserved image of early Germanic phonology and morphology (e.g. Loikala 1977: 229-230).

4.5. Baltic.

Latvians and Lithuanians speak languages classified within the Baltic languages of the Indo-European language family. Estonians, on the other hand, speak a Uralic language that falls within the Finnic branch. Laitinen et al. (2002) suggest, based on their assessment of the genetic data, that Latvians, Lithuanians and Estonians descended from a common population based on the similar frequencies of the N1a-M46 mutation in all three populations. They also support their conclusion by providing evidence from the archaeological record and by citing a Uralic relic found in the Latvian language. According to the study, this suggests that Estonians maintained their ancestral language, whereas Latvians and Lithuanians shifted languages, perhaps as the result of the Slavic expansion. Recent higher resolution data from Ilumae et al. (2016) confirms this hypothesis. As shown by Supplementary Table 14.9, N1a-VL29 represents about a third of genetic variation among Latvians, Lithuanians and Estonians. The same data also links Baltic populations with Samoyed populations in northern Siberia, and with the Finns and Sami of Scandinavia.

4.6. East Slavic.

Russians speak an Indo-European that falls within the East Slavic branch. Among Russians, the N1a-M46 mutation potentially signals the genetic legacy of populations that shifted from Uralic to Slavic about 1,500 years ago. It should be noted that most of the N1a-M46 variation among Russians appears to be N1a-Z1936 and N1a-VL29 (see Supplementary Tables 14.7 and 14.9). Additionally, among ethnic Russians the frequency of haplogroup N-M46 is influenced greatly by geography, with a diminishing north to south frequency cline. In northern Russia about forty-three percent of ethnic Russians have the mutation, whereas the frequency decreases to ten percent in the south (Balanovsky et al. 2008). This frequency pattern and the associated reindeer hypothesis are important components of a model that explains the origins of Slavic languages as a whole. Specifically, the shift to Slavic occurred in Eastern Europe without a large population expansion from a proto-Slavic homeland. It should be noted that another key component of the Slavic Expansion Model is downstream variants of the R1a-M420 mutation. Additional details will follow in Chapter 17: Section 9.

4.7. Altaic Languages.

For purposes of this present discussion, the term “Altaic” refers to potential areal relationships rather than a so-called “genetic” relationship for Turkic, Tungusic and Mongolic languages as advocated by some linguists. Supplementary Tables 14.19 and 14.20 suggest that the N1a-P43 and N1a-M46 mutations are useful markers for deciphering the linguistic prehistory of all three language families. These data from both tables also raise an interesting question, whether N1a-P43 and N1a-M46 are Paleolithic relics or, alternatively, if they define more recent population expansions associated with the *reindeer hypothesis*. N1a-M46 data for Altaic-speaking reindeer herders favor the more recent expansion (see Supplementary Table 14.13). Such a position is also supported by informative N1a-M46 variants for Altaic populations as a whole. At this point the reader is directed to Supplementary Table 14.8. As shown by the table, the N1a-M2019 mutation appears to be an important marker for exploring the genetic history of Tungusic and Turkic-speaking population in central Siberia. The N1a-F4205 represents an important marker among Mongolic speakers in East Asia (see Supplementary Table 14.11). Finally, the N1a-B479 mutation appears to be the genetic signature of Nanai people, a Tungusic-speaking population (see Supplementary Table 14.10).

As previously detailed in Chapter 6: Section 7, the C2-M217 mutation attains a significant frequency among many of the Altaic-speaking populations. C2-M217 should be seen as a Central Asian component among these populations, whereas N1a-M46 is the North

Eurasian component. The absence of C2-M217 within an Altaic population may suggest shift to an Altaic language by former Uralic-speaking reindeer herders. This appears to be the case among the Yakuts. Alternatively, Altaic-speaking populations with C2-M217 and N1a-M46 may suggest assimilation of a smaller Uralic-speaking group by a larger Altaic-speaking population. This assimilation may or may not have changed the subsistence strategy within the new admixed population. Tuvans and Buryats, for example, are two Altaic populations that have significant frequencies of C2-M217 and N1a-M46. Reindeer herding has been utilized by Tuvans whereas it appears that Buryats have never utilized this subsistence strategy (Mirov 1945).

4.8. Altaic and Transeurasian.

As previously detailed in Chapter 6: Section 7, the Transeurasian hypothesis has been formulated to explain striking lexical and grammatical similarities found among the Japonic, Koreanic, Turkic, Tungusic, and Mongolic language families. As explained in this section, the C2-M217 mutation offers a genetic tool to assess the possibility that all five languages shared a common prehistoric population. Downrange variants of the N-M231 haplogroup, as previously detailed, attain a significant frequency among some populations that speak Turkic, Tungusic, and Mongolic languages. Moreover, N-M231 is found in about four percent of Koreans (Park et al. 2012), and less than one percent of Japanese (Hammer et al. 2006; Sato et al. 2014). As such, N-M231 invites analysis that raises the possibility of a common ancestral population that spoke proto-Transeurasian.

Unfortunately, geneticists have not sequenced Koreans for informative downstream variants of the N-M231 haplogroup. The limited data from Hammer et al. (2006) suggests that most of the haplogroup N-M231 variation among the Japanese belongs to N1b-F2930, a mutation that is restricted to East Asia (see Section 2). As such, the available data currently fails to link Japanese and Koreans with the Altaic component of the Transeurasian Hypothesis.

4.9. Paleo-Siberian Languages.

The term “Paleo-Siberian” represents a convenient descriptor for several of the small North Eurasian language families including Yukaghir, Yeniseian, Eskimo-Aleut, and Chukotko-Kamchatkan as well as Nivkh language isolate. Interestingly, haplogroup N-M231 has not been detected among the Kets, a Yeniseian population for which data is available. Similarly, haplogroup N-M231 is absent among the Nivkh. (See Rootsi et al. 2007). However, N1a-M46 (Supplementary Table 14.20) attains a significant frequency among Koryaks and Chukchi, two East Siberian populations that speak a Chukotko-Kamchatkan language. Additionally, the same mutation is found among the Yupik, an Eskimo-Aleut population of the same region. Finally, the N1a-M46 mutation attains a significant frequency among Central Siberian Yukaghirs .

Yupik, Chukchi and Koryaks belong to N1a-B202, one of the highly resolved N1a-M46 variants reported by Ilumae et al. (2016). This observation provides additional support for the *reindeer hypothesis* with the idea that better reproductive success fueled a population expansion from north-central Siberia to a geographic dead-end at the Bering Sea. Further support for this idea stems from the observation that reindeer herding is practiced by the Chukchi and Koryaks. Unfortunately, Ilumae et al. (2016) were unable to identify the genetic history of the Yukaghir lower than N1a-P298 mutation. This suggests that their genetic history includes higher resolution downstream N1a-M46 mutations that have not been discovered.

Sometimes the unexpected absence of a haplogroup within a population presents useful data for researchers. Among the Yupik, who are sometimes referred to Siberian Eskimos, the frequency of haplogroup N-M231 hovers around fifty percent of the population (i.e. Ilumae et al. 2016). However, haplogroup N has been not found in North American Eskimos although they, like the Yupik, speak languages belonging to the Eskimo-Aleut language family. Founder effect and genetic drift may explain this observation. Alternatively, Paleo-Eskimos had already crossed over the Bering Sea before the expansion of haplogroup N-M231 into Eastern Siberia. Another explanation may stem from the paucity of Y-chromosome data for Native Alaskans.

Section 5. Conclusions.

The genetic, paleo-climatological and anthropological evidence suggest that NO-M214 evolved in Central Asia around 42 thousand years ago. N-M231 then evolved from NO-M214 about 38 thousand years ago in China. Diversification within N-M231 began close to the end of the last Ice Age with the evolution of N1a-F1206 and N1b-F2930. N1b-F2930 remained in East Asia. N1a-F1206 carries the story of genetic diversity in northern Eurasia.

N1a-M46 evolved roughly 13 thousand years ago, in northern China or Mongolia. Several informative downstream markers within N1a-M46 suggest a rapid bidirectional human expansion across northern Eurasia about five thousand years ago. Some researchers attribute the expansion to the development of metallurgy. However, the genetic, paleo-climatological and anthropological evidence suggest that this expansion resulted from better reproductive success. Thus, the rapid bidirectional human expansion across northern Eurasia conforms to the *reindeer hypothesis*. Moreover, the *reindeer hypothesis* not only explains a human expansion, but also the expansion of Uralic languages. In doing so, the hypothesis provides additional support for a global pattern of linguistic evolution that follows the *early farming dispersal hypothesis*. Finally, the reindeer hypothesis also confirms the role of language contact in shaping linguistic diversity in Scandinavia, the Baltic Region, Eastern Europe and Northern Eurasia. Nevertheless, it should be emphasized that hundreds of cultures and languages are dispersed across this vast landmass each with their own history. As such, the *reindeer hypothesis* hardly represents the final word for exploring the tremendous amount of cultural and linguistic diversity found in these regions. Rather, the hypothesis merely represents a starting point for future linguistic research that integrates archaeological, historical, genetic and linguistic perspectives.

Better resolution of language variation in Eurasia will require better resolution of the N1a-P43 mutation. The internal phylogeny of N1a-P43 still remains “unexplored territory.” N1a-M46, on the other hand, represents, by far, the most informative branch within N-M231 thanks to Ilumae et al. (2016) and the informative downstream mutations that they reported. Nevertheless, additional high-resolution sequencing of previously collected N1a-M46 samples would be fruitful. For example, high resolution N1a-M46 data are only available for thirty-nine Finns.

Chapter 15: Haplogroup O-M175.

Section 1. Overview of O-M175.

At this point the reader is invited to review Supplementary Figure 1.1 from the first chapter. As shown on the second page of the figure, the NO-M214 mutation is a downstream variant of KR-M526. Data previously presented in Chapter14: Section 2, suggests that NO-M214 arose roughly 42 thousand years ago in Central Eurasia. As such, the marker stands as a genetic relic of the colonization of the Eurasian landmass by modern humans via a so-called northern route. Then in northern China, or perhaps Central Asia, O-M175 and N-M231 evolved from NO-M214 roughly 38 thousand years ago.

Based on a synthesis of the data from Chapter 14: Section 2 and that provided by Gavashelishvili and Tarkhnishvili (2016), Paleolithic diversification within NO-M214 seems to follow a split in subsistence strategies. Some exploited the large game resources of the "Mammoth steppes." This explains why N-M231 is now found in Northern Eurasia. Meanwhile others found what they needed further south amongst the Paleolithic savannah and woodlands of what is now contemporary China. This partially explains how O-M175 became a significant marker for deciphering the population history of East Asia. Wang and Li (2013b) estimate, for example, that the haplogroup attains a frequency of about seventy-five percent among contemporary populations in China.

The reader is directed to Supplementary Table 15.1 which details the frequency and distribution of haplogroup O-M175 from a regional perspective. As shown by the table, O-M175 is not only an informative marker for East Asia, but also for South Asia, Island Southeast Asia, and Oceania. The reader is further invited to review Supplementary Figures 15.1 and 15.2 which provide a phylogenetic overview of O-M175 and its downstream variants. As shown by the figure, O-M175 has two main branches within its internal phylogeny, O1-F265 and O2-M122. Data from Poznik et al. (2016) suggest that this initial diversification of O-M175 variation occurred roughly 30 thousand years ago just before the Last Glacial Maximum. As detailed in the sections below (2-16), within O1-F265 the O1a-M119 mutation is an informative marker for deciphering the prehistory of Austronesian languages. O1b-M95 has emerged as an especially strong marker for Austro-Asiatic languages. The O1b-SRY465 mutation represents and especially informative marker for Japonic and Koreanic. Within O2-M122, O2a-002611 represents the genetic signature of Chinese languages. O2a-B451 is the genetic signature of Austronesian languages. The internal phylogeny of O2a-M134 awaits better resolution. Nevertheless, its downstream variants elucidate the prehistory of all the East Asian language families. Finally, the O2a-M7 mutation is an informative marker for Hmong Mien and Austro-Asiatic languages.

Section 2. Origins of East Asian Rice Cultivation.

A particularly striking observation of contemporary demography in South Asia, East Asia and Island Southeast Asia is high population density. This observation, of course, raises an important question. How can China, India, Korea, Japan, Bangladesh, and Malaysia, for example, support such large populations? The answer is rice. The story of rice cultivation in East Asia begins about 10 thousand years ago within the Yangtze River basin of eastern China (see Figure 15.1 below). Holocene climate change brought a regular pattern of monsoon rain

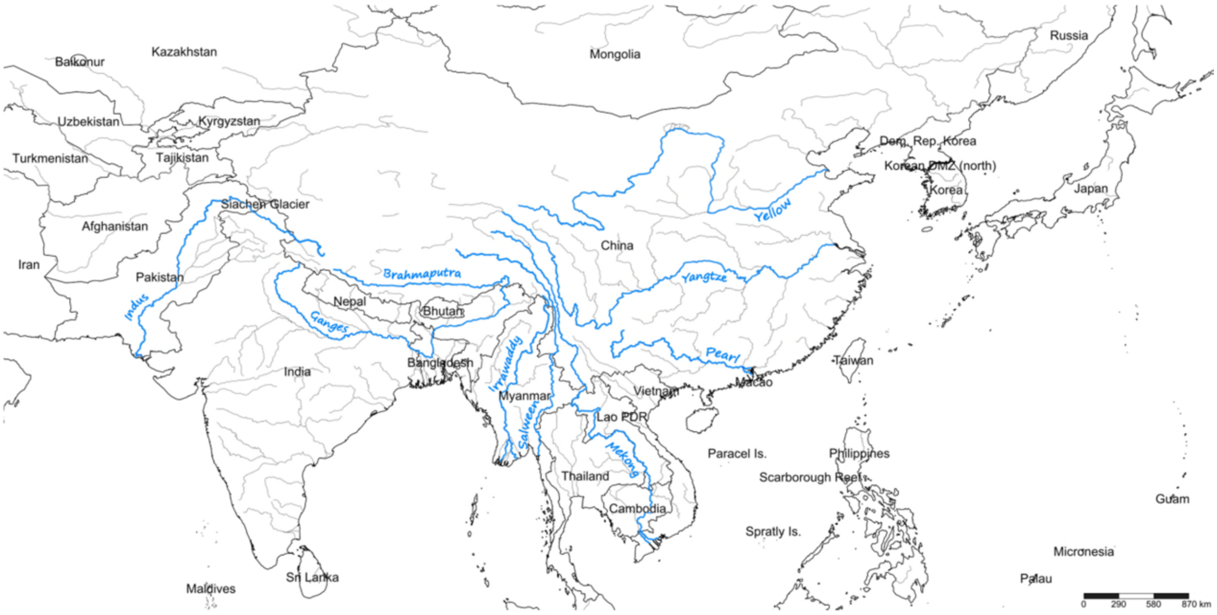


Figure 15.1. Rivers of East and South Asia.

to the region (Bellwood 2005: 111). The change in climate created conditions that are ideal for wet rice agriculture. According to Stevens and Fuller (2017), by around 4.5 thousand years ago rice cultivation began to spread out of the Yangtze River basin as the result of population pressure. Over a period of roughly two thousand years, rice cultivation spread into southern China, and then into Southeast Asia, which includes Thailand, Vietnam and Malaysia. Additionally, rice cultivation spread eastwards into Korea and Japan. Finally, Chinese rice cultivation spread westwards to India.

For linguists, the development of paddy field rice cultivation is significant. Languages thrive and survive because people have found a way to thrive and survive. The high population density associated with rice cultivation explains why several East Asian language families now occupy a large corner of the tapestry of global language variation.

Section 3. Early Expansion of Chinese.

The term "Chinese" requires additional clarification at this point. In term of ethnicity, the term "Chinese" refers to the Han ethnic group. The Han are found predominately in China. They are also one of fifty-six different ethnic groups recognized by the Chinese government. Today they comprise almost ninety-two percent of the population in China, and as such, they are the largest ethnic group in this country (e.g. CIA World Factbook 2019). From a linguistic perspective, "Chinese" represents one of two main branches within the Sino-Tibetan language family (*Ethnologue* 2018). The other branch is Tibeto-Burman. Within the Chinese branch, *Ethnologue* lists fourteen different languages. Furthermore, *Ethnologue* reports that around 1.3 billion people speak Chinese.

LaPolla (2001) provides an authoritative overview of the origins and expansion of Chinese languages from a linguistic perspective. First, he correlates the origins of Chinese with the evolution of rice cultivation in the Yangtze River basin. Then, he correlates the initial expansion of Chinese with the expansion of rice agriculture out of this region. As noted previously in Section 3, this occurred about 4.5 thousand years ago. Finally, LaPolla explains that the contemporary distribution of Chinese also reflects internal migrations and population displacements in China as well as the rise and fall of empires and kingdoms across East Asia.

Genetic evidence also supports a model of Chinese origins in the Yangtze River basin. Wang et al. (2013a) identify the O2a-002611 mutation as a potential signature marker of the Han expansion during the Neolithic. The study utilized a large data set of almost eight thousand samples, and as such the conclusions are especially compelling. Another study, Yan et al. (2014) takes that position that forty percent of Chinese can trace their ancestry to three "Neolithic super-grandfathers." According to the study, the genetic signature of these "super-grandfathers" is the O2a-F11, O2a-M117 and O2a-F114 mutations. It should be noted that O2a-F11 is a downrange variant of O2a-002611. Additionally, O2a-F114 and O2a-M117 are sister clades positioned downstream from O2a-M134 (see Supplementary Figure 15.2 for additional details).

At this point the reader is invited to review Supplementary Tables 15.2, 15.3 and 15.4. Based on the available data, O2a-002611 and O2a-F114 represent reliable markers for deciphering the prehistory of the Chinese language branch. The O2a-M117 mutation, on the other hand, represents not only an informative marker for Chinese, but also Tibeto-Burman, Austronesian, Austro-Asiatic, Tai-Kadai, Koreanic, and Japonic (see Sections 5-7 and 9-11 of this present chapter). However, the usefulness of this marker is very limited because informative downstream markers await identification. One significant problem stemming from the lack of informative downstream markers is attempting to determine when and where the O2a-M117 mutation evolved. Based on the available archaeological and genetic data, it appears that the mutation evolved about eight thousand years ago along the Yangtze River. The genetic data follow the contemporary distribution of O2a-M117 (see Supplementary Table 15.4), the dating estimate for O2a-F114 from Ning et al. (2016), the evolutionary history of O2a-002611 (Wang et al. 2013a), and the evolutionary history of O2a-N6 (Wei et al. 2017a). The archaeological data follow the history of rice cultivation in China (e.g. Stevens and Fuller 2017) and its expansion along this river system (e.g. Zhang and Hung 2008).

As noted earlier, Wang et al. (2013a) identified O2a-002611 as a potential signature marker of the Han expansion during the Neolithic. The study suggests that O2a-002611 arose about 12 thousand years ago in southeastern China. Later, during the early Holocene, people with the mutation migrated northwards to the Yellow and Yangtze River basins. This Paleolithic migration agrees with the archaeological data. Around the onset of the Holocene in China, Paleolithic hunter-gatherers congregated along the coastline of southeastern China (Zhang and Hung 2012). Then, during the Holocene, perhaps as the result of diminishing marine resources, some of these hunter-gatherers moved inland towards the Yellow and Yangtze rivers while other remained in southeastern China. Evidence for this scenario comes from contrasting tool making traditions and subsistence strategies among populations in southeastern and central eastern China during the late Paleolithic (Bellwood 2005: 126-127, Zhang and Hung 2012; Zhang and Hung 2013). Hunter-gatherers of the Yellow and Yangtze river basins are characterized by a unique micro-blade tool making tradition. Their tools differed from the pebble tool Hoabinhian tradition found in southeastern China and northern Vietnam. Additionally, the hunter-gatherers in east central China focused of the collection of tubers and nuts, whereas the Hoabinhian utilized aquatic resources.

These observed differences in food gathering and tool making traditions in late

Paleolithic China represent important data for linguists. Part of the story of linguistic diversity in East Asia seems to follow the Neolithic transition among two different Paleolithic cultural traditions. Chinese, Tibeto-Burman and Austro-Asiatic seem to follow the adoption of rice agriculture by the Paleolithic micro-blade cultures between the Yangtze and Yellow Rivers. Austronesian and Tai-Kadai seem to follow the adoption of rice agriculture by the Paleolithic pebble tool culture. This archaeological scenario potentially explains the genetic data. Dating estimates from Karmin et al (2015) potentially place the evolution of O2a-P164 in southeastern China about 20 thousand years ago. The O2a-P164 bifurcates into O2a-M134 and O2a-N6. Mutations downstream from O2a-M134 are potential genetic relics of the micro-blade culture, and O2a-N6 represents the pebble tool culture.

Section 4. The Expansion of Tibeto-Burman.

Tibeto-Burman languages were introduced in Chapter 4: Section 5 and the discussion of haplogroup D-M174. From a linguistic perspective, Tibeto-Burman languages are a branch within the Sino-Tibetan language family. According to Ethnologue (2018), the Tibeto-Burman branch consists of 442 languages that are organized within twelve different subbranches. Tibeto-Burman languages are predominately found in the East Asian countries of China and Myanmar (Burma), and the South Asian countries of India, Nepal, Bhutan and Bangladesh. A reliable estimate for the number of Tibeto-Burman speakers could not be found. The number is probably less than 100 million. From an archaeological perspective, the starting point for a discussion of this language group begins with the Tibetan Plateau in China. While rice cultivation explains the expansion of Chinese languages, the initial expansion of Tibeto-Burman correlates well with the cultivation of barley on the Tibetan plateau beginning about 3.6 thousand years ago. Unlike other grain crops, barley tolerates the cold and dry climate that is associated with the high altitude of this region (Zhang et al. 2016).

In addition to the cultivation of barley, an evolutionary adaptation also explains the success of Tibeto-Burman languages. The Tibetan Plateau lies at an average altitude of 4,000 meters above sea level. Here, hypoxia and altitude sickness pose a significant health danger. People from lower altitudes can, over time, become acclimated to living at high altitude. Nevertheless, Tibetans have an evolutionary adaptation that allows them to utilize the depleted oxygen level more efficiently than those who have moved to Tibetan from a lower altitude (Wu and Kayser 2006). A recent study (Yang et al. 2017) focuses on nine different sections of the human genome (or loci) that potentially control this evolutionary adaptation. The study compared the genomes of about three thousand Tibetans with seven thousand non-Tibetans from East Asia. This comparison indicates that the Tibetans and Han separated about 4.7 thousand years ago, which is consistent with the Y-chromosome and archaeological data.

As noted previously in Chapter 4: Section 4, a particularly significant genetic characteristic of Tibetans and the Tibetan Plateau is the elevated frequency of haplogroup D-M174. According to Qi et al. (2013) around fifty-four percent of Tibetans have the mutation. Additionally, thirty-three percent of Tibetans belong to O-M175, which is a common East Asian haplogroup. Qi et al. (2013) suggest that haplogroup O-M175 represents a Neolithic component of the Tibetan gene pool, a genetic relic of the westward expansion of agriculture into the region from central China. Haplogroup D-M174, on the other hand, reflects a much earlier hunter-gatherer component, a relic of the human colonization of East Asia.

According to Qi et al. (2013), the O2a-M117 mutation represents around ninety percent of the O-M175 variation among the Tibetans. As previously noted in Section 5, the O2a-M117 mutation stands as an especially strong genetic signature of the Han expansion, and with that, the expansion of the Chinese language branch. Taking this a step further, O2a-M117 seems to support the linguistic position that classifies Tibeto-Burman and Chinese as branches within the Sino-Tibetan language family. This follows linguistic and anthological interpretations of the data that posit a population split of proto-Chinese and Proto-Tibeto-Burman speakers on central plains of Yellow River valley about 6.5 thousand years ago (LaPolla 2013; Zhang et al. 2016). A similar position is also taken by the geneticists and their interpretation of the O2a-M117 data (Kang et al. 2012; Wang et al. 2014).

As stated earlier, Tibeto-Burman languages are also found in South Asia. It appears that these languages expanded into the region as the result of population pressure on the Tibetan Plateau. Unlike the Tibetan Plateau however, haplogroup D-M174 attains a low frequency among the Tibeto-Burman speaking populations of India, Nepal, Bhutan and Bangladesh (e.g. Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008, Gazi et al. 2013, Tamang et al. 2018). Rather, haplogroup O-M175 and the downstream O2a-M117 mutation point to Tibet as the source of Tibeto-Burman languages that are found in South Asia. Unfortunately, much of the South Asian data reports frequency results for poor resolution markers that are upstream from O2a-M117. One exception is Debnath et al (2011). They report O2a-M117 frequencies between twenty-five and forty-two percent for Tibeto-Burman-speaking populations in Eastern India. Gayden et al. (2007) is another exception. They report O2a-M117 frequencies between twenty-one and eighty-four percent for Tibeto-Burman-speaking populations in Nepal.

It should be noted that the Tibeto-Burman languages of Myanmar are classified within the Ngwi-Burmese sub-branch. Historical evidence suggests the expansion of Ngwi-Burmese from the Tibetan plateau is unrelated to the expansion of Tibeto-Burman sub-branches found in South Asia, such as Central Tibeto-Burman, Sal, and Western Tibeto-Burman. Rather than an agricultural expansion, Ngwi-Burmese correlates better with the rise and fall of the Pyu civilization and their migration along the Salween River (see La Polla 2013: 206-207).

Section 5. Trans-Himalayan.

Linguists should be aware of a proposed “Trans-Himalayan” language classification that recently appeared in a genetic study of South Asia (Tamang et al. 2018). Before discussing the concept, it should be emphasized that we defer to *Ethnologue*. This classification system generally reflects a natural division of a vast amount of linguistic, genetics, archaeological and historical data. As noted previously, *Ethnologue* classifies Tibeto-Burman and Chinese as the two main branches of the Sino-Tibetan language family. However, the linguist George van Driem contests this arrangement. He views Chinese (or Sinitic) as a group somewhere within Tibeto-Burman (Driem 2005). He also proposes a Trans-Himalayan phylum that contains many of the Tibeto-Burman languages (Driem 2014).

The Trans-Himalayan model is controversial (see LaPolla 2016). Nevertheless, the Trans-Himalayan discussion exposes a potential problem with the *Ethnologue* Sino-Tibetan classification. The Sino-Tibetan family classification seems to be an unnatural division of the data, especial from an anthropological perspective. The origins and expansion of Chinese languages stem from the cultivation of rice and the associated phenomenon of high population density. The origins and expansion of Tibeto-Burman languages, on the other hand, partly stems from the success of barley cultivation on the Tibetan Plateau. Another huge factor is genetic adaptations that allow Tibetans to thrive and survive at high altitude. Perhaps a more natural division of the data entails the creation of a Tibeto-Burman language family and a separate Chinese language family.

Section 6. Origins and Initial Southward Expansion of Austronesian.

The Austronesian language family occupies a large corner of the global linguistic tapestry with over twelve hundred languages and 324 million speakers (*Ethnologue* 2018). The Austronesian language family has two main branches, Formosan and Malayo-Polynesian. Formosan consists of twenty languages found on the island of Taiwan. Malayo-Polynesian, on the other hand, consists of 1236 languages that have a north to south geographic distribution from Taiwan to New Zealand, and a west to east distribution from Madagascar to Rapa Nui (Easter Island).

The Formosan branch of the Austronesian language family represents the linguistic signature of Taiwanese aboriginals as well as a linguistic relic of the prehistoric Dapenkeng culture. The Dapenkeng migrated to Taiwan from the East Asian mainland about 5.5 thousand years ago. For almost a thousand years the Dapenkeng were hunter-gatherers. Their subsistence strategy included the harvesting of marine resources. Then about 4.8 thousand years ago they adopted agriculture and began to cultivate foxtail millet and rice (Hung and Carson 2014). Shortly thereafter, about four thousand years ago, as the result of soil depletion and population pressure (Bellwood 2005: 135), rice agriculture spread from Taiwan to the Philippines. Linguistically, this expansion triggered a split in the Austronesian language family and the evolution of Malayo-Polynesian branch. From an archaeological perspective, the expansion follows the migration of the Lapita culture, a term derived from a unique style of pottery. The Lapita culture initially expanded southwards through the Philippines to Borneo. From Borneo, around 3.4 thousand years ago (Bellwood 2005: 137), a second Austronesian expansion occurred, with some migrating westwards in the direction of Malaysia, while others migrated eastwards in the direction of New Guinea.

From a genetics perspective, two different branches within the O-M175 phylogeny, are especially helpful for deciphering the initial expansion of Austronesian, O1a-M119 and O2a-N6. Within O1a-M119, the O1a-M307 and O1a-M110 mutations are the most informative. Within O2a-N6, the most informative mutation is O2a-B451. For additional information, the reader should review Supplementary Figures 15.1 and 15.2; Supplementary Tables 15.5, 15.6, and 15.7; Mirabal et al. (2012); Trejaut et al. (2014), and Wei et al. (2017a).

Section 7. Eastward Expansion of Austronesian into Oceania.

According to Bellwood (2005: 134-141) around three thousand years ago the Lapita culture began to spread across eastern Indonesia and Papua New Guinea (see also Horsburgh and McCoy 2017). By around two thousand years ago the Lapita culture reached western Oceania. Finally by around 1250 AD after colonizing many of the islands of central and eastern Oceania, the Lapita cultural expansion terminated in New Zealand. The genetic picture of this secondary expansion is rather interesting and complicated. In order to draw a simpler picture, one could compare the Austronesian expansion to a city bus with a long route that begins in Taiwan and ends in New Zealand. In Taiwan, passengers with the O1a-M307, O1a-M110 and O2a-B451 mutations started the journey. On New Guinea, passengers with C1-M208, M-256 and S-B254 climbed aboard the bus. In western Oceania, the passengers from Taiwan reached end their journey. The passengers from Papua New Guinea rode the bus to the end of the line.

The more complicated picture of the eastward Austronesian expansion points to the survival of cultural continuity despite population replacement. Such a conclusion is based on a comparison of data from several different sources. The reader is asked to review Supplementary Tables 15.5 and 15.8. Supplementary Table 15.5 reports for the O2a-B451 mutation. Supplementary Table 15.8 reports for all O1a-M119 mutations (i.e. O1a-M110 and O1a-M307) in Island Southeast Asia and Oceana. The reader is also invited to review the previous discussion of M-256 and S-B254 as provided in Chapter 13: Section 2 and the supporting information found in Supplementary Tables 13.1 and 13.2. Finally, the reader may want to review the discussion of the C1-M38 and C1-M208 mutations as presented in Chapter 6: Section 3.

The Austronesian colonization of eastern Indonesia and Papua New Guinea is reflected by the frequency pattern of O1a-M119 and O2a-B451 mutations in Island Southeast Asia. After the Austronesians arrived in this region, admixture occurred between this group and the Papuan-speaking populations. This is reflected by C1-M208, M-P256 and S-B254 mutations found in the Austronesian-speaking populations of the region and the fact that these mutations are the genetic signature of Papuan-speaking populations who had inhabited Island Southeast Asia for about 50 thousand years prior to the arrival of Austronesians. Finally, the genetic data suggest that a new population carried Austronesian eastwards from New Guinea into Oceania

According to Horsburgh and McCoy (2017) only a third of the Y-chromosome variation in Polynesia has a potential East Asian/Taiwanese origin. The remaining variation originated on New Guinea. Interestingly, based their analysis of mitochondrial DNA, which provides a maternal genetic perspective, the genetic history of Polynesian women is almost entirely linked to East Asia and Taiwan. This asymmetrical picture of population origins requires additional attention in the future. Perhaps this asymmetry simply reflects the effect of genetic drift and founder effect in Oceania. Such a conclusion is consistent with the data from New Guinea which reflect East Asian and Papuan female admixture. Kayser et al (2008), for example, reports data from the Admiralty Islands of Papua New Guinea. Around forty percent of the mtDNA has Papuan origins and the remainder has an East Asian origin. Another example comes from Delfin et al. (2012) and data for the Solomon Islands. The Papuan contribution is potentially twenty-two percent, and the remainder is East Asian.

Section 8. Expansion of Austronesian into Western Indonesia and Malaysia.

Previously in Chapter 13: Section 1, the concept of the so-called Wallace Line was discussed. The division was initially introduced to describe botanical features that are unique to Island Southeast Asia versus those unique to East Asia (see Blust 2013: 6-7 for additional information). See, also, Supplementary Figure 6.3 from Chapter 6). Today this term conveniently delineates western Indonesia from eastern Indonesia. Additionally, the Wallace line is important for anthropology as one finds human phenotype differences on both sides of the divide. Finally, a study from 2010 (Karafet et al.) found significant genetic differences between those living west of the division compared to those living on the eastern side. Haplogroups M-P256, S-B254, C-M38 and K-M526* represent an especially strong Papuan component of populations in eastern Indonesia. However, this component is weak or absent in western Indonesia. Populations in western Indonesia, on the other hand, have a strong East

Asian component, the O1b-M95 and O2a-M7 mutations. In eastern Indonesian, this East Asian component is essentially absent. Turning now to potential genetic signatures of an Austronesian component in Indonesia, it should be noted that data from Karafet et al. (2010) suggests that the Austronesian contribution hovers around fifty percent in western Indonesia, whereas the figure is around thirteen percent for eastern Indonesia. This conclusion was extrapolated from the frequency results for the O1a-M119, O1a-P203 (O1a-M307), O1a-M110, and O2a-P201 mutations.

The above discussion of the Wallace line helps to provide geographic and genetic context to a westward expansion of Austronesian languages from Borneo roughly 3.4 thousand years ago. Malaysians and Indonesians predominately speak Austronesian languages. However, Austro-Asiatic languages remain part of the linguistic diversity found in this area of the world. Moreover, Austro-Asiatic languages arrived in the region just before the Austronesians. The archeological record supports such a scenario. According to Bellwood (2005:139), rice agriculturalists from Thailand migrated onto the Malay Peninsula roughly 4.5 thousand years ago. Such a scenario is also consistent with genetic data. According to Arunkumar et al. (2015) the O1b-M95 mutation reflects a southward expansion of Austro-Asiatic-speaking rice farmers from Laos around this time. Data from Karafet et al. (2010) and Arunkumar et al. (2015) further suggest that after the Austronesians arrived, admixture occurred between this group and the Austro-Asiatic populations. Thus, part of the story of Austronesian languages in western Indonesia and Malaysia appears to entail language shift from Austro-Asiatic to Austronesian.

Section 9. Westward Expansion of Austronesian into East Africa.

Linguistic diversity in East Africa is generally associated with languages that fall within the Afro-Asiatic, Nilo-Saharan or Niger-Congo language families. Thus, it is rather unexpected to find Austronesian languages on the island of Madagascar. Linguistic, historical and genetic data provide an explanation.

Linguistic evidence and the Great Barito languages branch place the origins of Malagasy languages somewhere in the vicinity of Indonesia. Great Barito is a sub-branch of the Malayo-Polynesian branch of the Austronesian language family. Ethnologue (2018) lists thirty-five Great Barito languages. Twenty-three of these languages are found either in Malaysia, Indonesia or the Philippines. The remaining twelve languages are found thousands of kilometers away off the eastern coast of Africa. These twelve languages are classified within a Malagasy branch of the Great Barito language branch. One of the twelve Malagasy languages is spoken on the Island of Mayotte. The remaining eleven are spoken on the island of Madagascar.

According to the archaeological and historical records (e.g. Blench 2010), Madagascar was initially settled by hunter-gatherers from East Africa about four thousand years ago. Austronesian contacts with East Africa began around two thousand years with the arrival of ships from Malaysia. At times the Malays conducted raids. They also traded extensively with the East Africans, perhaps to obtain cinnamon. The East Africans, in turn, might have received chickens, bananas and taro root. Finally, around 1,500 years ago, after centuries of raids and trade, the Malays established permanent settlements on Madagascar.

For the purposes of discussing the genetic data, it is important to emphasize that the term “Malagasy” also has an ethnic connotation and describes the inhabitants of Madagascar.

According to the genetic data, the Malagasy people are a blend of populations from East Africa, the Middle East and Island Southeast Asia (Capredon et al. 2013; Poetsch et al. 2013; Tofanelli et al. 2009a). Based on the frequency of the O1a-M110 and O1b-M111 mutations, the Austronesian contribution among contemporary Malagasy is about twenty percent.

The Ma'anyan people represent a potential source population for the Malagasy language based on the linguistic evidence. An interesting study from 2015 (Kusuma et al.) explored a potential genetic connection between the Malagasy and the Ma'anyan people of Borneo. The study failed to find any evidence of a special genetic relationship between the two groups. Rather, the genetic evidence can only pinpoint Malagasy origins within a region, either Indonesia or Malaysia. Their findings seem consistent with the anthropological data provided by Blench (2010). He reports the Great Barito languages came from Indonesians who were pressed into service onboard the Malay ships. As such, it appears that the Austronesian settlers of Madagascar represented several different Indonesian and Malaysian ethnic groups that utilized a common Great Barito language as a *lingua franca*.

Section 10. The Austronesian Advantage.

Based on the number of speakers and its vast geographic distribution, Austronesian is indeed a significant linguistic "heavyweight." Donohue and Denham (2010) provide a useful summation of approaches and issues with respect to the history of Austronesian languages. They take a position that correlates the success of Austronesian with trade network that were controlled by Austronesian-speaking populations. These trade networks flourished because technological advantages, such as outrigger canoes. The also flourished because of good navigational skills that eventually carried the Austronesians over vast stretches of open water (for a more detailed discussion, see Blust 2013: 11-17).

Correlating the so-called success of Austronesian with trade is problematic because trade does not necessarily produce a reproductive advantage. In other words, Austronesian behaves much like language families that co-expanded with early agriculture, such as Niger-Congo or Sino-Tibetan. These languages thrived and survived because agriculture can support much higher population density than hunting and gathering. Bellwood (2005: 141) explains the importance of agriculture and its role in the Austronesian expansion. He writes that most Austronesian-speaking populations practice agriculture and without it, the Austronesians could not have colonized Oceania. In short, agriculture appears to be a far more crucial component of the Austronesian success story than just trade networks and technology.

As explained earlier in Section 6, the Austronesian agricultural expansion began with rice and millet cultivation on Taiwan. As explained by Bellwood (2005: 130-139) and Blust (2013: 6-7), when the Austronesian expansion reached Borneo, climatic conditions no longer supported the cultivation of grain crops. At this point the Austronesians began to cultivate tubers and tree crops that flourish in Island Southeast Asia and Oceania. Tubers include taro and yams. Examples of tree crops are sugar cane, bananas, pandanus, breadfruit, sago palm, canarium nuts and coconuts.

Nevertheless, correlating the success of Austronesian with tuber and tree crop agriculture also seems problematic. Papuans also cultivate these crops (see Chapter 13: Section 4). Why, then, would agriculture have been hugely successful for Austronesians, and moderately successful for Papuans?

As previously detailed in Chapter 13: Section 4, during the Holocene the Papuans of New Guinea congregated in the central highlands of this island. Around ten thousand years ago, Papuan agriculture began with the cultivation of bananas and sugar cane. Coastal lowland agriculture on New Guinea occurred much later, about three thousand years ago with the arrival of Austronesians.

Malaria could explain the differing Papuan and Austronesian patterns of early agricultural activity on New Guinea. During the Holocene the Papuans may have moved to the highlands to avoid malaria. Austronesians, on the other hand, were able to farm the coastal areas of the island because of an evolutionary adaptation that made them resistant to tropical splenomegaly syndrome, a massive and fatal enlargement of the spleen that occurs as the result of chronic exposure to malaria. It should be noted that researchers recognize twelve epidemiological zones of malaria where people have endured chronic long-term exposure to the affliction. The expansion of Austronesian falls within Malaysian and Australasian epidemiological zones. Combined, both zones include the Philippines, Indonesia, the Malay Peninsula, Indonesia, West Timor and Papua New Guinea (Arrow and Gelbrand 2004: 142-143 and Table 6-1).

Evidence for the so-called "Austronesian advantage" stems from a study published by Clark and Kelly in 1993. The researchers compared gamma globulin polymorphisms from Austronesian and non-Austronesian populations on New Guinea. Gamma globulin was examined because the marker has a strong association with the immune system. The researchers were able to identify a specific polymorphism characteristic of lowland Austronesian-speaking populations who are resistant to tropical splenomegaly syndrome. They also identified another polymorphism associated with highland Papuan groups, populations that are highly susceptible to tropical splenomegaly syndrome.

Clark and Kelley made several points that are useful for understanding the Austronesian advantage. Anopheles mosquitos are the "vector" that transmits *Phasmodium*, the parasite that causes Malaria (see Cox 2010 for more details). These mosquitos thrive in the wet and swampy lowlands of New Guinea, whereas they are far less prevalent in the highlands. Furthermore, lowland coastal agriculture furthers intensifies the spread of malaria by creating habitat that facilitates the breeding cycle of these mosquitos. Finally, lowland coastal agriculture creates permanent human settlements which provide a host population for the *Phasmodium* parasite. The "Austronesian advantage" suggests that Austronesians could farm the coastal areas of New Guinea, whereas such activity for Papuans would have been lethal.

Since the prevalence of malaria diminishes with altitude, malaria avoidance would conveniently explain why the Papuans occupied New Guinean highlands at the onset of the Holocene. This presupposes the presence of malaria on New Guinea before the arrival of the Austronesians. Such an argument seems plausible based on a recent study of the *Plasmodium vivax* organism, a species of the *Plasmodium* parasite that is especially prevalent in the so-called Austronesian world. Loy et al. (2018) report a close genetic relationship between *Plasmodium vivax* parasites that infect chimps and gorillas in Africa and the *Plasmodium vivax* parasites that infect people in Island Southeast Asia. *Plasmodium vivax* parasites from both regions share a common ancestor that evolved in Africa. When people migrated out of Africa around 100 thousand years ago, the *Plasmodium vivax* organism essentially "hitched-a-ride" with the humans.

Clark and Kelley (1993) also suggest that admixture between Austronesians and Papuans created a new population that inherited a genetic resistance to tropical splenomegaly syndrome. Consistent with recent Y-chromosome evidence, their data suggest that some Papuans eventually joined the lowland Austronesian farming settlements on New Guinea. Papuans and Austronesians and then produced children. The children inherited an evolutionary adaptation to tropical splenomegaly syndrome as well as an Austronesian language.

The above discussion of the so-called "Austronesian advantage" serves a linguistic purpose. The contemporary distribution of Austronesian languages reflects that enhanced reproductive success mediated the early expansion of this language family. This is an important point because greater reproductive success stands as an essential characteristic of Bellwood's early farming dispersal hypothesis (Bellwood 2005: 1-11). However, as noted earlier, correlating agriculture with greater reproductive success seems problematic for Austronesian when one considers that Papuans also practiced a similar form of agriculture as the Austronesians. The "Austronesian advantage" resolves this discrepancy.

Section 11. Austro-Asiatic.

Based on data from *Ethnologue* (2018), the Austro-Asiatic language family consists of 167 languages spoken by around 105 million people. These languages are distributed across South and East Asia, primarily in India, Bangladesh, Myanmar, Cambodia, Laos, Thailand, Vietnam and Malaysia. Within the Austro-Asiatic language family the Munda and Mon-Khmer branches form the two main divisions. Munda represents the Austro-Asiatic languages of South Asia and Mon-Khmer represents the Austro-Asiatic language of East Asia. Among the Austro-Asiatic language, Vietnamese has attained official language status in Vietnam and Khmer is the official language of Cambodia.

Efforts to identify the putative homeland of Austro-Asiatic languages have produced three different models that place the origins of this language family either in India, the Mekong River Valley of Laos, or southern China. In order to discuss the Mekong River Valley model, it should be noted that the Mekong is among the major waterways of East Asia. It flows over four thousand kilometers from the Tibetan Plateau through Yunnan province into Myanmar, Laos, Thailand, Cambodia and Vietnam, where it empties into the South China. The Mekong River Valley is located where the borders of Myanmar, Laos and Thailand converge on map. The linguist Paul Sidwell (2010) has identified this area as the geographic point of origin for Austroasiatic languages. He favors the Mekong River Valley with the idea that the region of greatest linguistic diversity also defines the geographic origins of a language family. Sidwell builds his argument through an analysis of the morphological, phonological, and lexical data.

George van Driem in a paper he published in 2011 advocates the India model of Austro-Asiatic origins based on phonological reconstructions. According to Driem, Austro-Asiatic originated in India because linguistic reconstructions point to a hot and humid tropical climate not found in southern China. Driem also offers botanical evidence. He suggests that modern domesticated rice originates from a hybrid of three different Neolithic variants, indica, japonica, and dry upland rice. He asserts that an initial hybrid of indica and dry upland could have only occurred somewhere near the Bay of Bengal. However, this suggestion conflicts with a more authoritative analysis. Fuller (2012) suggests that modern domesticated rice is a hybrid of just two variants, proto-indica from India and domesticated japonica from China.

The southern China model identifies the Three Gorges Region of the Yangtze River as the putative homeland of Austro-Asiatic languages. Higham (2002), based on his interpretation of the archaeological and linguistic evidence, argues that Munda and Mon-Khmer split about six thousand years in this region of Sichuan province. Austro-Asiatic languages and rice agriculture then expanded upstream along the Yangtze into the Yunnan province of southwestern China. Austro-Asiatic and rice agriculturalists then expanded out of Yunnan along major river systems. According to Higham (2002), river systems were utilized to avoid travel through the dense forest canopy. Munda and rice agriculture migrated into northeastern India along the Brahmaputra river. Mon-Khmer and rice agriculture radiated southward from Yunnan along several major river systems including the Mekong, Irrawaddy, Chao Praya and Red.

Higham (2002) provides linguistic support for this model by offering proto-Austro-Asiatic reconstructions for terminology related to rice agriculture. He also cites Mon-Khmer languages found in China, especially those that fall within the Palaungric sub-branch. One of these languages is the U (or P'uman) which is spoken by the Blang ethnic group in the in Yunnan province. According to Higham (2002), U has the distinction of being the northernmost Austro-Asiatic language. Furthermore, its location on the Mekong River support a close correlation between the expansion of early rice agriculture and the expansion of Mon-Khmer languages via this waterway.

The southern China model is also supported by archaeological and botanical data that explain the Munda/Mon-Khmer split in the Austro-Asiatic family. Focusing now on Munda, Zhang and Hung (2008) date the arrival of rice agriculture in Yunnan at around four thousand years ago. Fuller (2012) suggests that shortly after the arrival of rice agriculture in Yunnan, japonica rice advanced westward into India along the Brahmaputra river and eventually onto the Ganges plain. It should be noted that the cultivation of proto-indica and other rice strains in India predate the arrival of japonica. However, pre-japonica rice agriculture in India was characterized by the casual dry land cultivation of a grain that was rotated with other crops. As such, pre-japonica rice agriculture never became a significant source of food in India (Fuller 2012; Bates and Singh 2017). Rather, rice only became a food staple about three thousand years ago, about a thousand years after the introduction of the japonica variety from China. By this time, japonica and indica had been developed into a high-yield hybrid rice. Furthermore, by this time farmers in India had perfected rice paddy cultivation, a technique that fully exploits the potential of rice agriculture. Thus, the arrival of Munda languages and japonica rice in India substantially altered the demographic landscape with the introduction of a rice variety that sustains high population density.

Turning now to the genetic evidence, the O1b-M95 mutation attains a significant frequency among the Munda and Mon-Khmer populations of Asia (see Supplementary Table 15.9). As such it represents a strong genetic relic of the Austro-Asiatic expansion (e.g. Chaubey et al. 2011). Additionally, the marker provides genetic support for the southern China model. Initial analysis of the O1b-M95 mutation has focused on the origin and age of the mutation to determine if the mutation expanded from East Asia to India, or alternatively, if it expanded in the opposite direction, from India to East Asia. Kumar et al. (2007) suggest that the mutation evolved about 65 thousand years ago among Mundari populations of India. Based on this interpretation of the data, they suggest that O1b-M95 eventually spread eastwards from India into East Asia. This conclusion, of course, has serious flaws including the fact that modern humans where not in South Asia 65 thousand years ago.

A recent study (Arunkumar et al. 2015), one that utilizes a substantial amount of data, suggests that O1b-M95 originated in Laos. Then about seven thousand years ago the mutation expanded out of the region is a star-like pattern. This analysis of the data seems to support the Mekong River Valley model of Austro-Asiatic origins, as advocated by Sidwell (2010). Nevertheless, the Arunkumar et al. 2015 study is problematic in that the research fails to identify the age of the O1b-M95 mutation. Rather, they focus on when the mutation expanded.

Another large data study from 2015 (Zhang et al.) also explored the evolutionary history of the O1b-M95 mutation. The researchers determined that mutation arose between 20 and 40 thousand years ago in southern China. The mutation then expanded into central China about 16 thousand years ago. This was followed by a westward expansion into India about 10 thousand ago. Ancient DNA data supports the idea that O1b-M95 populations migrated into India along the Yangtze River. Li et al. (2007) report O1b-M95 mutations from Neolithic remains found along this waterway. Taking this a step further, the Yangtze River was the corridor carried Austroasiatic languages into Yunnan province and beyond.

Unfortunately, researchers have mostly relied upon short tandem repeat markers to interpret the evolutionary history of the O1b-M95 marker. The preferred and most reliable method entails the identification of informative single polymorphisms that are downstream from O1b-M95. Besides a lack of downstream mutations, another huge problem with interpreting the evolutionary history of O1b-M95 is the lack of data for Myanmar, Vietnam and Malaysia. Nevertheless, a conservative analysis of the available genetic data suggests that the O1b-M95 mutation has Paleolithic origins. Taking this a step further, the weight of genetic, archaeological and linguistic evidence (Higham 2002; Li et al. 2007; Zhang and Hung 2008); Zhao 2011; Fuller 2012; Zhang et al. 2015) favors the southern China model of the Austro-Asiatic expansion. Specifically, about four to five thousand years ago, Austro-Asiatic-speaking rice farmers arrived in Yunnan, China. From this region, proto-Munda and rice agriculture expanded westwards into India. Proto Mon-Khmer and rice agriculture expanded southwards in the direction of Myanmar, Thailand, Laos, Cambodia, Vietnam and Malaysia.

Section 12. Austronesian and Tai-Kadai.

According to *Ethnologue* (2018), about 81 million people speak a Tai-Kadai language. This language family consists of 91 different languages and is found in China, Thailand, Vietnam, Laos and Myanmar. A handful of Tai-Kadai languages are also found in India. The Tai-Kadai language family has three main internal branches: Hlai, Kam-Tai and Kra. Hlai consists of two languages found on Hainan Island in China. Kra has sixteen languages that are found in China and Vietnam. The remaining 72 languages belong to the Kam-Tai branch. Finally, one of the Tai-Kadai languages, Thai, with around 60 million speakers, attains official language status in Thailand.

Linguists generally point to southern China as the putative homeland of Tai-Kadai languages (e.g. Sidwell 2013). From an archaeological perspective, Zhang and Hung (2012) raise the possibility that Tai-Kadai represents a linguistic relic of hunter-gatherers who eventually adopted farming and rice cultivation. A recent genetic study (Brunelli et al. 2017) takes the same position. Blench (2013) notes, however, that the prehistory of Tai-Kadai remains somewhat murky. He suggests that the historical kingdom of Siam best explains the position attained by Tai-Kadai languages within the contemporary tapestry of global language variation.

Although the genetic, linguistic and archeological evidence point to mainland China as the putative homeland of Tai-Kadai, a genetic study suggests Hainan Island. As noted earlier, the Hlai branch of Tai-Kadai consists of two languages that spoken on the island. Li et al. (2008) take the position that Hlai is a linguistic relic of the earliest Tai-Kai languages and that it evolved on Hainan Island among the aboriginal populations. Their arguments stem primarily from analysis of the elevated frequency of O1a-M119 and O1b-M95 mutations among the Hainan aboriginals. Nevertheless, the assertion made by Li et al. (2008) requires more archaeological evidence as well as the identification of higher resolution polymorphisms downstream from O1a-M119 and O1b-M95. From a linguistic perspective, it seems just as plausible that Hlai evolved from a Tai-Kadai language once spoken in mainland China (Norquest 2007). Then, at some point, speakers of early Hlai made a sea crossing to the island.

The weight of archaeological, linguistic and genetic data identifies Taiwan as the putative homeland of Austronesian languages (see Section 6). As such, in situ evolution of Austronesian on Taiwan is potentially undermined by linguistic discussions of Austronesian influences found in Tai-Kadai languages. For example, Sagart (2004) takes the position that Tai-Kadai is a sub-branch within the Austronesian language family. Thurgood (1994), on the other hand, offers compelling arguments for a borrowing relationship between Austronesian and Tai-Kadai somewhere in Guizhou and Guangxi provinces about four thousand years ago.

Those searching for common prehistoric ancestral language for Tai-Kadai and Austronesian on mainland Asia may find support from the archaeological record. As previously detailed in Section 4, a pebble tool Hoabinhian cultural tradition evolved in southeastern China and northern Vietnam. Additionally, genetic data may provide further support. The O2a-N6 mutation could be especially informative marker (see Wei et al. 2017a). However, an alternative explanation would associate Austronesian influence on Tai-Kai with the historical rise and fall of the Austronesian-speaking Champa civilization, which Sidwell (2013) dates between 500 BC and 1500 AD. Such a position was taken by Doi (2012) and his discussion of Austronesian influences found in Austro-Asiatic languages, especially Vietnamese. An extension of Doi's argument suggests that an intense language contact relationship not only existed between the Champa and Austro-Asiatic populations, but perhaps Champa and Tai-Kadai populations.

Focusing now on the genetic evidence for Champa influence, Li et al. (2013) report Y-chromosome data for the Utsat people who are a potential relic population. Their results suggest that the Austronesian expansion onto mainland East Asia was carried by a small population from Island Southeast Asia. This small Austronesian-speaking population then admixed with a larger non-Austronesian-speaking mainland population. A similar conclusion was reached by He et al. (2012a) in their study of ethnic Cham and Kinh in Vietnam as well as Thais and Laotians.

Section 13. Hmong-Mien and Language Contact.

The contemporary distribution of Hmong-Mien languages is found in the mountain regions of southern China, northern Laos and northern Vietnam. According to *Ethnologue* (2018), the Hmong-Mien family consists of thirty-nine languages and 9.3 million speakers.

This language family has three main divisions: Hmongic, Mien and Ho Hte. Hmongic consists of thirty-three languages, Mien of five languages, and Ho Hte of one (the She language). Among the linguists, one finds consensus for a putative Hmong-Mien homeland in Southern China (e.g. Benedict 1987; Kosaka 2002; Driem 2011). The main controversary among linguists is the relationship between Hmong-Mien and Tai-Kadai, Austro-Asiatic, and Tibeto-Burman (see Ostapirat 2018 for an overview). Some linguists look for macro-family relationships, and others favor language contact.

From an anthropological perspective, the temporal starting point for a discussion of the Hmong Mien language family is the arrival of rice cultivation in southeastern China during the Neolithic (e.g. Zhang and Hung 2010). A study from 2011 (Cai et al.) provides an interesting genetics perspective concerning the origins of the Hmong-Mien family. The researchers observed significant frequencies of O1b-M95, O2a-M7, and O2a-M117 among Austro-Asiatic and Hmong-Mien-speaking populations. Based on their analysis of these data, the researchers suggest a common origin language model for both language families. Nevertheless, an alternate interpretation of the genetic data suggests that similarities found in both languages were shaped by the convergence of speech communities and outcomes as predicted by language contact theory. Based on the data (see Supplementary Table 15.4), the O2a-M117 mutation represents a useful marker for deciphering potential language contact between Hmong-Mien, Austro-Asiatic, Tai-Kadai and Tibeto-Burman. O1b-M95, on the other hand, helps to decipher language contact between Hmong-Mien, Austro-Asiatic, and Tai-Kadai (see Supplementary Table 15.9). Finally, the O2a-M7 mutation points to the language contact between Hmong-Mien and Austro-Asiatic (see Supplementary Table 15.10).

Section 14. Koreanic.

14.1. Overview of the Linguistic Data.

Around 77 million people speak Korean (*Ethnologue* 2018). This language is largely confined to the Korean Peninsula in East Asia. As previously discussed in Chapter 6: Section 10, the classification of the Korean language has proven difficult. The so-called "southern theory" attempts to associate Korean with Dravidian or Austronesian. The "northern theory" classifies Korean as part of an Altaic macro-family. *Ethnologue* (2018) classifies Korean as one of two languages within the Koreanic family. Finally, one could argue that Korean is more characteristic of language isolate rather than a language family. As such, the potential Altaic, Chinese and Austronesian influences in Koreanic are the result of language contact. In order to evaluate these models of Koreanic origins, the discussion that follows will evaluate archeological and genetic data from the Korean peninsula.

14.2 Overview of the Archaeological Record.

It should be noted that the Paleolithic period on the Korean Peninsula is poorly documented within the archaeological record. As such the Neolithic triggers a temporal starting point for deciphering the prehistory of Koreanic (see Kim 2015). In order to discuss the Korean Neolithic, we must turn to the origins of agriculture in China. One center of early agriculture was the Yangtze River basin and the evolution of rice cultivation (see Section 3). Another center of early agriculture in China is located along the Yellow River between the Mongolian steppes and Huai River (Zhao 2011). Here, roughly eight thousand years ago, the Xinglonggou culture began to cultivate foxtail and broomcorn millet. About two thousand years later, millet cultivation expanded from this region of northeastern China to the Korean Peninsula (e.g. Stevens and Fuller 2017).

The expansion of millet agriculture onto the Korean Peninsula signals the beginning of the Neolithic within this region. Archaeological discussions often refer to the Korean Neolithic as the Chulmun period, a term that describes a unique form of pottery. Although the arrival of early agricultural often triggered rapid population growth in many areas of the world, the situation in Korea appears to be different. The archaeological record fails to support rapid population growth during the Chulmun period. According to Ahn (2010), millet merely supplemented a hunter-gatherer diet. Instead, it was the arrival rice cultivation from China that triggered the rapid population growth that is characteristic of early agricultural expansions, and with that, the transition from foraging to farming.

Rice cultivation came to Korea roughly 3.5 thousand years ago during the Korean Bronze Age, or the so-called Mumun period, a term that also describes a unique form of pottery. Archaeologists (e.g. Ahn 2010) have identified three potential source regions from which this expansion may have occurred. Potentially, rice may have expanded onto the Korean Peninsula from southeastern China, somewhere near the Pearl River delta. Alternatively, rice could have expanded from central China somewhere near the Yangtze River delta. Nevertheless, most archaeologists favor northeastern China and more specifically, the Shandong and Liaodong Peninsulas. This area conveniently avoids a sea crossing and suggests that Chinese farmers migrated onto the Korean Peninsula because rice cultivation failed in Manchuria due to climatic conditions.

14.3. Overview of the Genetic Data.

At this point the reader may want to review Chapter 6: Section 7 for an overview of C2-M217 mutations and the discussion of the *Transeurasian hypothesis*. The reader may also want to review Supplementary Figure 6.4 and the overview of the internal phylogeny of C2-M217. As reflected by the figure, C2-M407 mutation is an informative C2-M217 variant. The C2-M407 mutation provides the strongest argument for a Mongolic contribution to the Korean gene pool. Based on frequency data provided by Kwon et al. (2015), the mutation attains a frequency of about thirteen percent among contemporary Koreans. Huang et al. (2017) identify Mongolia as the source of these mutations. The study further suggests that C2-M407 evolved about 5 thousand years ago. Another study, Zhong et al. (2010) suggests 12 thousand years ago with an expansion about 10 thousand years ago. Based on these dating estimates, C2-M407 variation in Korea stands as a potential relic of the Neolithic expansion of millet cultivation from northeastern China.

As noted in the previous paragraph, about thirteen percent of Koreans have a variant of the C-M130 main haplogroup. However, the highest frequency main haplogroup is O-M175, which represents eighty percent of the population (Kwon et al. 2015). Among the Koreans, the most informative downstream variants of O-M175 are O2a-002611, O2a-M117, O2a-F114, and O1b-SRY465.

14.4. O2a-002611 among Koreans.

It should be noted that the O2a-002611 mutation was previously discussed in Section 4. This mutation represents a particularly strong marker of the Chinese Neolithic and the emergence of the Han ethnic group. Among contemporary Koreans, this mutation attains a frequency of around ten percent (Kwon et al. 2015). The source of the mutation is probably eastern central China.

14.5. O2a-M117 and O2a-F114 among Koreans.

At this point the reader is invited to review Supplementary Figure 15.2. The O2a-M134 mutation has two downstream variants, O2a-M117 and O2a-F114. Among contemporary Koreans the O2a-M117 mutation attains a frequency of around thirteen percent and the O2a-F114 mutations attains a frequency around ten percent (see Supplementary Tables 15.3 and 15.4). The source of O2a-F114 among Koreans is eastern central China.

Turning now to O2a-M117, this mutation is found throughout East Asia. It should be noted that the O2a-M117 mutation represents an important marker for deciphering the prehistory of Chinese (see Section 4) and Tibeto-Burman (see Section 5). Moreover, it is an important marker for deciphering language contact between the Hmong-Mien, Tai-Kadai, Austro-Asiatic and Tibeto-Burman language families (see Section 9).

Unfortunately, linguistically informative downstream variants of O2a-M117 remain unknown. A possible exception is the O2a-M133 mutation, which represents almost all the O2a-M117 variation among Koreans (see Park et al. 2012). Interestingly, the O2a-M133 mutation has also been detected among the Han Chinese on Taiwan as well as among some of the aboriginal Austronesian-speaking aboriginal populations on the island (see Supplementary Table 15.11). As the reader may recall, the Taiwanese aboriginals are descendants of the Dapenkeng culture that migrated to the island from mainland Asia about 5.5 thousand years ago. The Han Chinese, on the other hand, migrated to Taiwan within the last four hundred years ago (e.g. Williams 2003). Thus, the source O2a-M133 mutations among contemporary Koreans requires additional research. It might be mainland China, or it might be Taiwan.

14.6. Koreanic and Austronesian.

Kim (2009) discussed potential Austronesian influences found in the Korean language. The presence of O2a-M133 mutation among Taiwanese aboriginals and Koreans seems to provide additional support for this position. Nevertheless, Kim (2009) also acknowledges that a relationship between Korean and Austronesian runs against mainstream opinion among contemporary linguists. Lee and Ramsey (2011: 27-28) further suggest that a relationship between Austronesian and Korean stems from Japanese linguistic research in the early twentieth century and efforts to undermine a sense of ethnic identity among Koreans.

The genetic, linguistic and archaeological data support the evolution of O2a-M117 somewhere on central plains of Yellow River. Here proto-Chinese and Proto-Tibeto-Burman separated. O2a-M117 eventually migrated westwards with speakers of early Tibeto-Burman onto the Tibetan Plateau (see Section 5). From Tibet, O2a-M117 expanded southwards into India, Burma and beyond (see Section 9).

Based on data from Trejaut et al. (2014), the O2a-M133 mutation dates to around nine thousand years for the Taiwanese aboriginals and thirteen thousand years for the Taiwanese Han. Additionally, Ning et al. (2016) date O2a-F114 to around eight thousand years. Since O2a-F114 and O2a-M117 are sister clades, it would follow that O2a-M117 also evolved eight thousand years ago. Thus, based on this limited amount of available data, it seems as though that O2a-M117 not only expanded westwards along the Yellow River, but eastwards as well. The eastward expansion terminated at the Yellow Sea. One population then followed the East Asian coastline northwards to Korea and another group migrated southwards to Taiwan. Such a scenario is supported by the frequency of O2a-F114 among Han Chinese and Koreans (see Supplementary Table 15.3). Additionally, it seems significant that O2a-M117 and O2a-F114 are found among Mongolic and Tungusic-speaking populations (see Supplementary Tables 15.3 and 15.4).

The discussion of the O2a-M117, O2a-M133 and O2a-F114 mutations certainly underscores a need to further explore the downstream phylogeny of O2a-M134. Such an effort could potentially identify more informative markers that helps to decipher language variation in East Asia. Additionally, such an inquiry could help to further clarify the geographic origins and expansion of the mutation. In short, the frequency of O2a-M133 among Taiwanese aboriginals and Koreans seems perplexing. Just as perplexing is the phylogenetically close O2a-B451 mutation and its frequency among Austronesian-speaking populations.

14.7. O1b-SRY465 among Koreans.

Based on data from Kwon et al. (2015), roughly one third of Koreans have the O1b-SRY465 mutation. The same study identifies two variants of the O1b-SRY465 mutation among Koreans: O1b-47z and O1b-L682. Based on the available data, O1b-L682 appears to have evolved in Korea. O-SRY-465 and O1b-47z, on the other hand, appear to have elsewhere as both mutations are scattered throughout East Asia (see Supplementary Tables 15.12 and 15.13). Kim et al. (2011) suggests that O1b-SRY465 evolved in northeastern China between six and ten thousand years ago. The same study dates O1b-47z variants in Korea at around four thousand years. Since O1b-SRY465 is scattered throughout East Asia, it seems as though the coastal expansion of this mutation closely follows that of O2a-M117 and O2a-F114.

14.8. Conclusions for Section 14.

The *Transeurasian hypothesis* was previously introduced in Chapter 6: Section 7. It represents a recent effort to classify Koreanic as part of a macro-language family. According to the hypothesis, Tungusic, Mongolic, Turkic, Koreanic and Japonic all trace their origins to a common proto-Transeurasian language. An alternative view of Koreanic would view this language family as a “near isolate.” As such, non-Koreanic influences are explained by the convergence of different speech communities. Interestingly, the available archeological and genetic data fail to advance one model over another. Rather, interpretation of the linguistic data seems to be the decisive factor. At the end of the day, the classification of Koreanic remains a highly subjective decision that entails consensus among the linguists. The *Ethnologue* (2018) classification is probably consistent with the opinion of most linguists.

Section 15. Japanese and Koreanic.

The internal linguistic phylogeny of the Japonic family, as well as the geographic distribution of Japonic language-speakers, seems to suggest that Japonic behaves much like a “near isolate” like Koreanic. Japonic has two main branches, Japanese with a single language, and Ryukyuan with eleven languages. The Japanese branch has, by far, the largest number of speakers, around 129 million. While the number of Ryukyuan speakers is unknown (Shimoji 2010), data from *Ethnologue* (2018) suggests around two thousand speakers. In terms of geographic distribution, Japanese is spoken throughout the entire range of the Japanese islands. Ryukyuan languages, on the other hand, are confined to the Ryukyuan Islands at the southernmost tip of the Japanese archipelago.

In terms of geography, contemporary Japan consists of a chain of islands that extend roughly three thousand kilometers north to south. As mentioned previously in Chapter 4: Section 4 and Chapter 6: Section 3, the gene pool of contemporary Japan has a very strong Paleolithic component, roughly forty percent. The genetic relics of the Paleolithic founding populations of Japan are the C1-M8 and D1b-M55 mutations. This genetic data is consistent with the archeological record which suggests that modern humans colonized the Japanese Islands roughly 30 thousand years. Around 16 thousand years ago their descendants evolved into the Jomon hunter-gatherer culture (e.g. Hudson 2013). The term “Jomon” describes a unique style of pottery that has become a signature relic these people.

Roughly two thousand years ago the so-called Yayoi people migrated from Korea to the Japanese island of Kyushu. They introduced agriculture and rice cultivation eventually replaced foraging as the main subsistence strategy in Japan. The O1b-SRY465 and O1b-47z mutations are the genetic relics of the Yayoi migration (e.g. Hammer et al. 2006). According to Naitoh et al. (2013) and Sato et al. (2014), roughly a third of Japanese belong to O1b-SRY465 and its downstream variants. As such the genetic evidence potentially supports linguistic arguments that posit a common ancestral language for Koreanic and Japonic (e.g. Whiteman 2012) or that Japonic could belong to a Transeurasian macro-family (see Chapter 6: Section 7). Nevertheless, some linguists question this relationship (e.g. Tranter 2012). Rather, Japonic and Koreanic are essentially seen as language isolates that influenced each other as the result of language contact and the expansion of agriculture from the Korean Peninsula to the Japanese islands.

As noted previously, efforts to build a linguistic macro-family relationship for Japonic and Koreanic are controversial. However, language contact influence exerted by Chinese in both language families is incontrovertible. For example, roughly half the lexicon in both language families has a Chinese origin (Kim 2009; Shibatani 2009). The influence of Chinese extends, in fact, to the earliest attestations of Koreanic and Japonic and efforts to adapt Chinese character script for writing Old Korean and Old Japanese (e.g. Tranter 2012). Given the indisputable influence that Chinese has played in shaping the Korean and Japanese languages, it should not be surprising to see a strong Han Chinese signature in the contemporary Japanese and Korean gene pools. For example, roughly ten percent of Koreans (see Section 10) and five percent of Japanese (Sato et al. 2014) have the O2a-002611 Chinese signature mutation.

Building a macro-relationship for Japonic and Koreanic seems problematic because it ignores massive Chinese language influence in both families and the potential contribution to Japonic from the Jomon hunter-gatherers. Like Koreanic, an alternative view of Japonic would certainly view this language family as a “near isolate.” Like Koreanic, the available archeological and genetic data fail to advance a macro-family model over a language isolate model for Japonic. Once again, interpretation of the linguistic data seems to be the decisive factor. Like that for Koreanic, the *Ethnologue* (2018) classification for Japonic is probably consistent with the opinion of most linguists.

Section 16. Japanese and Austronesian.

According to Naitoh et al. (2013) and Sato et al. (2014) eight to nine percent of Japanese have the O2a-M134 mutation. As previously detailed in Section 10, better resolution of O2a-M134 variation among Koreans may clarify the extent of Austronesian language contact with speakers of early-Koreanic languages. Additionally, clarification of O2a-M134 mutation among Japanese may also clarify the extent of Austronesian language contact with speakers of early Japonic. Interestingly Robbeets (2017b) recently presented linguistic arguments for language contact between early Japonic and early Austronesian speech communities. The researcher favors a "para-Austronesian" speech community on the Shandong peninsula of northeastern Chinese mainland. She also places the putative homeland of Japonic languages on the Chinese mainland and the Liaodong peninsulas. According to the researcher, both speech communities converged roughly four thousand years ago.

Another possible model for explaining potential Austronesian and Japanese would posit an Austronesian expansion from Taiwan onto the southern Ryukyuan islands beginning about 4.5 thousand years ago. Archeological support for the Ryukyuan model comes from tool and pottery remains found on the Japanese Yaeyama Islands, which are located at the southernmost tip of Ryukyuan atoll, about 250 kilometers east of Taiwan. Two reports (Summerhayes and Anderson 2009; Hudson 2017) identify these items as possible Austronesian artifacts from Taiwan. Additionally, both reports raise the possibility of contact between Jomon hunter-gatherers and Austronesians in the vicinity of Okinawa Island. While the available archeological evidence fails to support an Austronesian migration onto Okinawa or points beyond, better resolution of O2a-M117 and O2a-N6 variation among East Asians may well paint a different story. Alternatively, trade relationships in the Taiwan Straits and along the Ryukyuan Islands may have resulted in the convergence of speech communities.

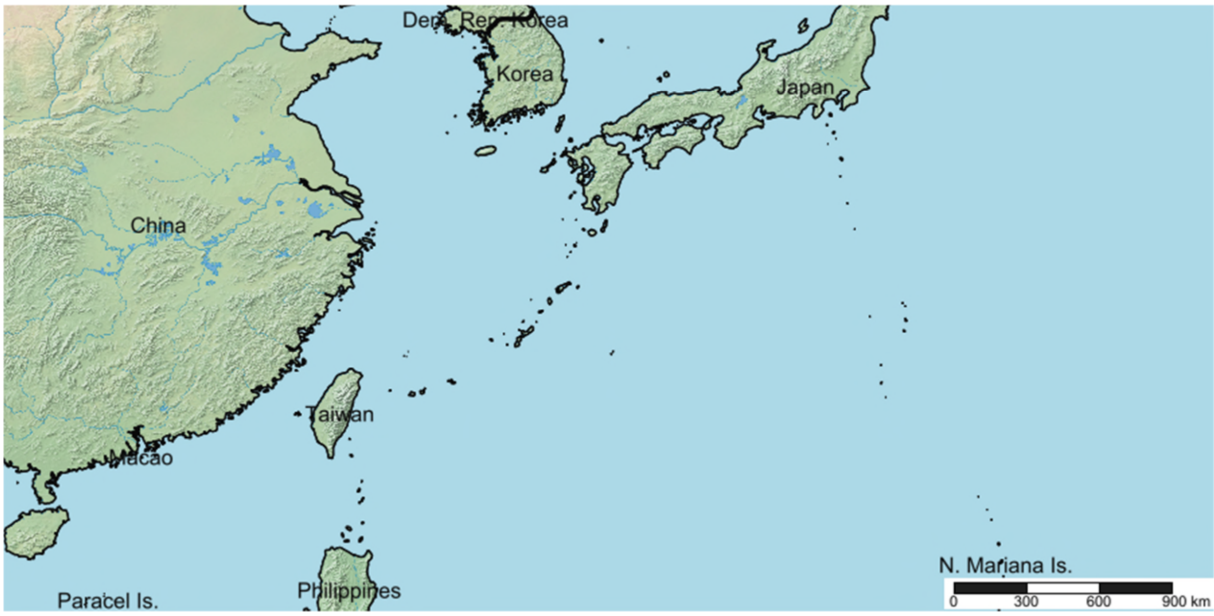


Figure 15.2. Mainland China, Taiwan, Japan and Korea.

Section 17. Conclusions.

The O-M175 main haplogroup has become a useful marker for understanding linguistic diversity in East Asia, South Asia, Island Southeast Asia and Oceania. This marker along with archeological and linguistic data suggest that early rice cultivation partially explains why Chinese, Austro-Asiatic, Austronesian, Koreanic, and Japonic now occupy a large corner of the tapestry of global language variation.

According to genetic and archaeological data, the expansion of early Chinese is associated with early rice cultivation in the Yangtze River basin. The origins and expansion of Tibeto-Burman languages, on the other hand, partly stems from the success of barley cultivation on the Tibetan Plateau. Another huge factor is genetic adaptations that allow Tibetans to thrive and survive at high altitude. Thus, the standard Sino-Tibetan language family classification seems to be an unnatural division of the anthropological data.

Austronesian languages probably evolved *in situ* on Taiwan. The eastward Austronesian expansion into Oceania points to the survival of cultural continuity despite population replacement. Austronesian-speaking populations are resistant to tropical splenomegaly syndrome, a condition caused by chronic exposure to malaria. As such, they can farm the malaria infested coastal regions of Island Southeast Asia.

The prehistory of Austronesian languages in western Indonesia and Malaysia appears to entail language shift from Austro-Asiatic to Austronesian. The Austronesian settlers of Madagascar represented several different Indonesian and Malaysian ethnic groups that utilized a common Great Barito language as a *lingua franca*.

From archaeological and genetic perspectives, the Neolithic transition on the Korean Peninsula triggers the temporal starting point for deciphering the complex prehistory of Koreanic languages. The available archeological and genetic data fail to advance one model of Koreanic origins over another. Rather, interpretation of the linguistic data seems to be the decisive factor. Similarly, interpretation of the linguistic data seems to be the decisive factor in defining the linguistic relationship between Japonic and Koreanic. Finally, a possible model for explaining potential Austronesian influences in Japanese might entail an Austronesian expansion from Taiwan onto the southern Ryukyuan islands beginning about 4.5 thousand years ago.

Chapter 16: Haplogroup Q-M242.

Section 1. Overview.

Although the Q-M242 haplogroup evolved in Asia, its downstream mutations have become an important tool for deciphering indigenous linguistic variation in the Western Hemisphere. We begin this overview by inviting the reader to review the second page of Supplementary Figure 1.1 from the first chapter. The P1-M45 mutation is positioned downstream from the KR-M526 paragroup. According to Poznik et al. (2016: Supplementary Table 10), the P1-M45 mutation evolved roughly 44 thousand years ago. We suggest that this occurred in southwestern Asia. Diverging from P1-M45 are Haplogroups Q-M242 and R-M207. The Q-M242 mutation evolved roughly 30 thousand years ago at the beginning of Marine Isotope Stage 2 (Poznik et al. 2016). This probably occurred in south-central Siberia.

The reader is now directed to Supplementary Table 16.1 which reports contemporary data. As shown by the table, the Q-M242 mutation is distributed throughout Eurasia where it represents a small percentage of the genetic diversity in this region. The reader should now review Supplementary Table 16.2 which reports Q-M242 data for contemporary indigenous populations of the Americas. As shown by the table, the Q-M242 haplogroup represent almost all the indigenous genetic diversity among Native Americans.

At this point the reader is invited to examine Supplementary Figure 16.1: Part A. The Q-M242 haplogroup has two main internal divisions within its phylogeny: Q1-F903 and Q2-L275. Both mutations evolved about 28 thousand years (Poznik et al. 2016). This probably occurred in south-central Siberia. Q2-275 mutations are confined to Eurasia and represent a very small fraction of the genetic diversity within this region (e.g. Huang et al. 2018). Furthermore, Q2-L275 mutations are not linguistically informative. Rather, linguistically informative mutations are downstream from Q1-F903.

Turning now to the diversification of Q1-F903 in Asia, as shown by Supplementary Figure 16.1: Part A, downstream from the Q1-F903 mutation is the Q1a-F1096 and Q1b-M346 mutations. Both mutations evolved about 23 thousand years ago (Wei et al. 2018). Two mutations diverge from Q-F1096: Q1a-F746 and Q1a-M25. Wei et al. (2018) suggest the split occurred 15 thousand years ago. As discussed below in Sections 7 and 8, the Q1a-F746 mutation helps to deciphering the prehistory of the Eskimo-Aleut language family.

Continuing with the diversification of Q1-F903, the Q1b-M346 mutation also diverges from Q1-F903. Downstream from Q1b-M346 is the Q1b-L54 mutation. According to Wei et al. (2018), this mutation evolved roughly 17 thousand years ago in south-central Siberia. Two linguistically informative lineages evolve from Q1b-L54: Q1b-M3 and Q1b-Z781. Both lineages evolved roughly 14.5 thousand years ago (Wei et al. 2018). These mutations are important for linguists as they represent the beginning of in situ genetic diversification of Q-M242 mutations in the Americas, and with that, the co-evolution of genetic, linguistic and cultural diversity among Native Americans (see Section 5).

Section 2. Quantity and Quality of Q-M242 Data among Native Americans.

Linguistic diversity among Native American is complex. However, genetic tools for deciphering this diversity are limited both in terms of the amount of data and the availability of informative mutations. Among Native American populations, the Q-M242 and C-M130 haplogroups define the indigenous genetic component. In North America, Q-M242 mutations represent about ninety-three percent of indigenous Y-chromosome diversity (e.g. Zegura et al. 2004). The remaining seven percent of Native American genes belong to C2b-P39. In Central and South America, Q-M242 represents almost all of the indigenous Y-chromosome diversity (e.g. Geppert et al. 2011; Roewer et al. 2013).

Underhill et al. reported the discovery of the Q1b-M3 Y-chromosome mutation in 1996. Despite this early success, the search for informative Native American mutations has proved elusive for geneticists. Much of the genetic data has only been published in the last two years (e.g. Wei et al. 2018; Grugni et al. 2019). Part of the problem with identifying informative Native American mutations may well stem from post-Columbian factors that reduced Y-chromosome variation among Native Americans. Shortly after the arrival of European in the fifteenth century, the Native Americans experienced a rapid decrease in population size. Many of them succumbed to European diseases against which they had not developed immunity. Furthermore, marriages between European men and Native American women may have reduced Y-chromosome genetic variation among the indigenous populations of the New World (for additional details, see Malhi et al. 2008).

Pre-Columbian factors may also explain the limited genetic variation found in Native Americans. Humans colonized the Americas relatively late in the game, about sixteen thousand years ago. Thus, genetic variation may be a question of time depth. Populations in Africa, for example, have diversified for around 300 thousand years.

A demographic model that surfaced in two studies (Regueiro et al. 2013; Roewer et al. 2013) may also explain reduced genetic variation among Native Americans. Both studies suggest that populations in Africa and Eurasia have a significantly different demographic history than populations in the Americas. Compared to Africa and Eurasia, the Americas never experienced a massive expansion of genetic variation that is characteristic of agriculture expansions. Rather, the human colonization of the Western Hemisphere occurred relatively quickly. This was then followed by isolation and tribalization of human populations, which continued until the arrival of the Europeans.

Battaglia et al. (2013) suggest that the extreme isolation found in the Americas may reflect a preference for hunter-gathering over agriculture as the primary survival strategy, even in areas where crops were cultivated. Bellwood (2005: 146-149) suggest that agriculture was limited in the Americans partly because of climate, partly because of the limited number of animals available for domestication, and partly because the only cereal crop was maize. He also questions the extent to which agriculture was a significant part of the survival strategy among those who cultivated crops. For example, maize, potatoes, and manioc are a potential source of calories. However, condiment crops, such as chilies and avocados, and the growing of squash for drinking gourds, are not staples.

Interestingly, language typology seems to reflect an inverse correlation between linguistic and genetic variation in the Americas. In other words, the characteristic leveling of linguistic diversity that accompanies agriculture expansion may not have occurred in the New World. Compared to Africa, Eurasia, and Oceania, linguistic diversity in the Americas appears much more diverse and more difficult to classify. *Ethnologue* (2016) lists a total of eighty-two language isolates for the world, and of these languages, sixty are found in the Americas. Furthermore, of the sixty-two unclassified languages listed by *Ethnologue* (2016), thirty are found in the Western Hemisphere, and more specifically, in South America (for more information, see Supplementary Table 16.3).

Despite the difficulties listed above, it is important to emphasize that the effort to identify informative Y-chromosome mutations among Native Americans is finally gaining momentum, especially in the last two years. However, more data is needed, especially for North America. Unfortunately, many of the North American native groups refuse to participate in genetic studies because of their historical mistrust of Europeans (e.g. Reardon 2017). Mulligan and Szathmary (2017) also suggest that Native Americans feel disrespected by researchers, and for this reason, they refuse to participate in genetic studies. Some researchers attempt to identify an Asian source population for Native Americans. However, contrary to what is circulated by these researchers, Native Americans take the position that they came from the Americas. To suggest otherwise is offensive.

Section 3. A Robust Model of Native American Origins.

3.1. Overview.

Section 3 offers an alternative model of Native American origins by detailing the beginning and the end of the Pleistocene mammoth hunting tradition. In doing so, this section departs from standard genetic and archeological models that define Asia as the geographic origin of Native Americans. Rather, arguments are presented defines Native Americans origins as a cultural tradition that evolved in the Americas. Such a position seems more consistent with archaeological, climatological, genetic and linguistic data.

3.2. Marine Isotope Stage 3 Expansions in Northern Eurasia.

Mammuthus primigenius, commonly known as the woolly mammoth, evolved roughly 450 thousand years ago. During the Last Ice Age, it achieved an astonishing range from Europe to the Americas (Kahlke 2015). At the beginning of Marine Isotope Stage 3, between 47 and 32 thousand years ago, *Homo sapiens* expanded across Northern Eurasia (e.g. Hamilton and Buchanan 2010). During this expansion human populations began to hunt the woolly mammoths that proliferated in this region. Archaeological support comes from the Sopochnaya Karga meteorological station which is located above the Arctic Circle in Siberia. Near the station researchers discovered the remains of a woolly mammoth that died 45 thousand years ago. Examination of the remains indicates that humans killed and butchered the animal (Pitulko et al. 2016).

Additional archeological evidence comes from the Sunghir archeological site located about 190 km northeast of Moscow. The remains of five males were found. They died roughly 34 thousand years ago. Evidence from the site further suggests that they hunted mammoths. One of the remains, the so-called Sunghir-1 man, was between 35 and 45 years old at the time of his death, which may have been the result of a hunting accident. Those that buried the man appear to have conducted a ritual. Sunghir-1 was buried with valuable stone tools. Valuable ornaments were also found. Sown onto his burial garments, for example, were thousands of mammoth ivory beads. Finally, his corpse was also covered in red ochre (for additional information, see Sikora et al. 2017: Supplementary Materials).

Important archaeological evidence also comes from the Yana Rhinoceros Horn site located above the Arctic Circle, where the Yana River empties into the Arctic Ocean. Archeological remains suggest that the site was used by Paleolithic mammoth hunters (Pitulko et al. 2004). Dental remains were recovered from two children who died here roughly 32 thousand years ago. Researchers determined that they have the P1-M45 mutation

(Sikora et al. 2018).

The archeological evidence presented above reflects that humans managed to adapt to the cold climate of Siberia during Marine Isotope Stage (MIS) 3. Part of this adaptation required specialized hunting skills needed to harvest a woolly mammoth, an animal that is about the same size as a modern-day African elephant. A gruesome discussion of the tactics utilized by Paleolithic hunters is provided by Pitulko et al. (2016) and one thing seems obvious - it must have been a dangerous undertaking. Nevertheless, the reward must have outweighed the risk. Successful prehistoric human adaptation to cold climate required a reliable high energy food supply, adequate clothing and shelter, raw materials for making cutting tools and projectile points, and fuel for fire. The woolly mammoth solved all these problems (Pitulko and Nikolskiy 2012; Pitulko et al. 2016; Pfeifer et al. 2019). A single woolly mammoth provides thousands of kilograms of meat which could be cached and stored long term in sub-zero conditions. Mammoth ivory produces exceptional projectile points. The hide can be used for clothing and tents. Dung and bones provide fuel for fire. Of course, cold adapted Paleolithic people in northern Eurasia ate other animals. Nevertheless, mammoth remains provide especially robust archaeological evidence that points to cultural continuity that had begun roughly 45 thousand years ago in Northern Eurasia and terminated roughly 11 thousand years ago with the onset of the Holocene.

3.3. Marine Isotope Stage 2 Hiatus.

The Last Glacial Maximum occurred about 27 thousand years ago (Clark et al. 2009). At this stage of the Last Ice Age, glaciation had reached its southern most extent in the Northern Hemisphere. While glaciation in Siberia was not as extreme as that in Europe, the region was nevertheless cold, arid and uninhabitable (e.g. Serdyuk 2005). Most of the region became depopulated during the Last Glacial Maximum. The human populations that once hunted above the Arctic Circle retreated below the 50th parallel, probably to the Altai region, which is located roughly where present-day Russia, China, Kazakhstan and Mongolia converge. They sought refuge in this area because, for a variety of reasons, the region was protected from glaciation and adverse weather conditions characteristic of the Last Glacial Maximum (Binney et al. 2017).

Hamilton and Buchanan (2010) characterize this retreat of population into the Altai region as a “hiatus” that lasted between 32 thousand and 16 thousand years ago. Human genetic and linguistic variation can be tied to the Ice Age hiatus with the idea that refugia consisted of reproductively isolating populations of humans (see, for example, Stewart and Stringer 2012; Gavashelishvili and Tarkhnishvili 2016). Among the human genetic diversity that evolved during the Ice Age hiatus in southern Siberia was haplogroup Q-M242, which diverged from P1-M45 roughly 30 thousand years ago (see discussion in Section 1). Archeological and genetic support comes from the banks of the Belaya River in southern Siberia and the remains of the Mal’ta boy. Raghavan et al. (2014) report that he died about 24 thousand years ago, roughly at the time of the Last Glacial Maximum. The researchers further report Y-chromosome data that place the child somewhere near the root of the R-M207 haplogroup, where haplogroups R-M207 and Q-M242 diverge from P1-M45.

3.4. Marine Isotope Stage 2 and Re-Expansion of the Mammoth Hunters.

Based on archaeological evidence, Hamilton and Buchanan (2010) time the end of the

Marine Isotope Stage 2 hiatus at 16 thousand years ago. The ice glaciers began to melt. Human populations across Northern Eurasia emerged from the southern refugia and expanded northwards. Within this region, weather conditions were characterized by widely oscillating warm and cool phases (Serdyuk 2005). For the purposes of this discussion, weather conditions become important because data reported by two different studies (Pitulko and Nikolskiy 2012; Mann et al. 2015) suggest that woolly mammoth populations in Siberia contracted and expanded rapidly based on the availability of forage. The availability of forage was tied to the warming and cooling cycles.

Genetic, archeological, botanical and climatological evidence support a rapid co-expansion of humans and woolly mammoths that began in southern Siberia about 16 thousand years ago and ended in Alaska about 14 thousand years ago. Pitulko and Nikolskiy (2012) and Mann et al. (2015) time the beginning of the co-expansion with the Bølling-Allerød warming phase. They report that warmer weather produced an abundance of forage for mammoths. The sudden increase in forage produced a sudden increase in mammoth populations. Population pressure then forced the mammoths to migrate across northeastern Siberia into Alaska in search of more food.

When the mammoth population increased, human populations increased. When the mammoths migrated, the humans followed. Dating estimates from Wei et al. (2018), about 17 thousand years ago, time the evolution of Q1-L54 to a gradual increase in air temperatures that preceded the Bølling-Allerød warming phase (see Figure 16.1 below). The study further suggests that southern Siberian origins of Q1b-L54 are supported by the current distribution of its downstream variants, Q1b-L330, Q1b-M3 and Q1b-Z781. Q1b-L330 remained in southern Siberia and became a predominant lineage in the Altai, Tuva, and Ket people of Siberia. Q1b-M3 and Q1b-Z781 evolved roughly 15 thousand years ago and mark the beginning of New World Q-M242 diversity.

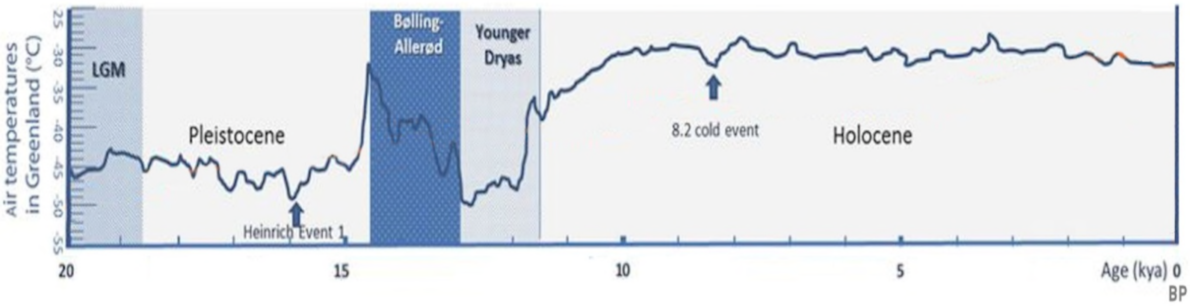


Figure 16.1. Bølling-Allerød Interstadial.

Original file was licensed under a Creative Commons Attribution 4.0 International License. See Platt, D. et al. 2017. "Mapping Post-Glacial expansions: the peopling of Southwest Asia. Scientific Reports 7**:** 40338. Link to File Image was modified and re-licensed under the Creative Commons Attribution-Share Alike 3.0 Unported license. Link to File

Dating estimates for Q1b-M3 and Q1b-Z781 correlate well with archaeological data from Swan Point, which is located in the Tanana Valley about 100 km southeast of Fairbanks, Alaska (Holmes 2011). Based on carbon-14 data obtained from carbonized grease and fat remains, humans occupied this site roughly 14 thousand years ago. Bones suggest mammoth and horses were on the menu. Additionally, evidence suggests that people used bones as fuel, a common practice for those living in tundra regions. Finally, their tools are similar to those from northeastern Asia.

3.5. The Mammoth Hunters of North America.

The oscillating pattern of cooling and warming periods during the Late Pleistocene ended with the Younger Dryas, a brief cold snap that lasted about 800 years, between 12.9 and 11.7 thousand years ago. The end of the Younger Dryas marks the beginning of the Holocene, and with that, higher temperatures and much more stable weather conditions. Warmer weather caused the ice glaciers to melt, which raised sea levels. The Bering land bridge, a 2,500-kilometer corridor that once connected northeastern Asia and Alaska, eventually disappeared underneath the rising sea.

The gradual geographic isolation of New World populations from those in Asia is recorded on the human Y-chromosome by downstream variants of Q1b-L54 mutations. According to Wei et al. (2018), New World diversification of Q1b-L54 began roughly 15 thousand years ago. For example, Anzick-1, the remains of a boy who died about 13 thousand years ago in Western Montana, has a mutation that is downstream from Q1b-Z780. The oldest Q1b-M3 remains come from Prince of Wales Island in Alaska and the Shuka Kaa man who died about 10 thousand years ago (Kemp et al. 2007).

The beginning of the Holocene also marks the end of a cultural tradition. After the mammoth hunters reached eastern Alaska, about 15 thousand years ago, the Cordilleran and Laurentide ice sheets may have blocked further migration southwards into contemporary British Columbia. At some point the process of deglaciation produced an ice-free corridor that allowed the hunters to continue the journey from Alaska to the Great Plains of North America (Dyke 2004; Potter et al. 2018). The archaeological record suggests that the journey through the ice-free corridor occurred by around 14,000 years ago. This figure is derived from dating estimates taken from a North American mastodon, a proboscidean closely related to the mammoth. The remains were uncovered at the Manis archaeological site in Washington State (Waters 2011). Researchers determined that the mastodon was killed by humans because a bone or antler projectile was found imbedded in one of the rib bones. Additional evidence for timing the passage through the ice-free corridor comes from dating estimates for the Anzick-1 child that found at a burial site in Western Montana. As noted previously, he died about 13 thousand years ago.

The Anzick boy represents a significant find in that he can be associated with the so-called Clovis culture. The term Clovis describes a unique type of projectile point and these points were found with Anzick-1. Traditionally, the Clovis culture is interpreted as the initial Native American cultural tradition. An alternative interpretation suggests that Clovis represents the end of the Pleistocene mammoth hunting tradition that began about 45 thousand years ago in Northern Eurasia. Frison (1998) suggests that mammoth hunting required a robust projectile point and Clovis points were very much up to the task. As such, Clovis points reflect a Paleolithic innovation in North America. This innovation perfected a subsistence strategy that had begun thousands of years earlier in Northern Eurasia. To underscore the idea of cultural continuity, we also note that the burial ritual of one of the first Pleistocene mammoth hunters, Sunghir-1, is remarkably similar to that of one of the last Pleistocene mammoth hunters, Anzick-1 (cf. Rasmussen et al. 2014; Sikora et al. 2017). Both were covered in red ochre. Both were buried with valuable tools.

According to Frison (1998) mastodon and mammoth hunting continued in North America until about 11 thousand years ago. Then, the mammoth suddenly disappeared. Perhaps they became extinct because of warmer weather, or perhaps it was human population pressure, or perhaps a combination of both (Fiedel 2008). Regardless of the reasons, the demise of mammoth hunting marks the beginning of the Native American cultural tradition.

Section 4. Classification of Native American Languages.

The above model of Native American origins provides a time component for building models of indigenous language variation in the Western Hemisphere. Now comes the question of information management or how the data should be organized to construct these models. A tripartite division of the data into Amerind, Eyak-Athabaskan and Eskimo-Aleut seems to work. This division follows Greenberg (1987) and his view on linguistic diversity in the Americas. Since Greenberg's classification is controversial among the linguists, an explanation is in order.

American Indian Languages, a reference guide published in 1997 by Lyle Campbell, is an authoritative must-have resource for linguists. The guide examines Native American languages from the perspective of historical linguists with the goal of classifying these languages. According to Campbell, the methodology employed by Greenberg is unconventional and unreliable. Campbell asserts that a rigorous application of the comparative method fails to prove the unity of Amerind as suggested by Greenberg.

Campbell's classification of Native Americans languages is consistent with contemporary mainstream opinion (e.g. *Ethnologue* or *Glottolog*). However, in the last thirty years Greenberg has often surfaced in genetic studies as an authoritative classification for Native American languages. So frequent was the use of Greenberg's classification that Bolnick et al. (2004) felt compelled to warn geneticists. In an article published in a science journal they labeled Greenberg's classification as problematic and non-standard. Nevertheless, Greenberg still surfaces in the genetic studies (e.g. Flegontov et al. 2016a).

Campbell's and Greenberg's approach to classification illustrates a strange dichotomy between a single discipline historical linguistic approach to Native American languages and an approach that integrates multidisciplinary perspectives. The historical linguistic approach clearly sides with Campbell. However, Greenberg provides a natural division of the data when multidisciplinary perspectives are employed. From a genetics perspective, Amerinds represent *in situ d*iversification of Q1b-L54 mutations in the Western Hemisphere. The genetic history of Eskimo-Aleut and Eyak-Athabaskan speakers, on the other hand, is potentially shaped by later geneflow across the Bering Sea. Culturally, Amerind reflects *in situ* adaptation to climate change in the New World. Eskimo-Aleut culture was hugely influenced by the development of active whaling which began in Japan roughly five thousand years ago. Linguistically, Amerind represents *in situ* diversification of languages in the Western Hemisphere since the beginning of the Holocene. Athabaskan and Eskimo-Aleut languages were potentially shaped by later cultural exchange with the indigenous peoples of northeastern Asia.

Our use of Greenberg's classification should not be interpreted as a disagreement with the contemporary standard classification of Native American languages (e.g. Campbell 1997; *Ethnologue* 2019; *Glottolog* 4.0). Rather, the Eskimo-Aleut and Eyak-Athabaskan language families require extra attention.

Section 5. Amerinds.

5.1. Overview.

Coinciding with the initial diffusion of Q1b-M3 and Q1b-Z780 mutations in the Americas was the beginning of the Holocene about 11 thousand years ago. The Bering land bridge became submerge under the rising sea level. The mammoths and other mega faunal resources disappeared. Former mammoth hunters now exploited new food resources found in the regions where they had settled. The earliest example of this transition comes from the Upper Sun River site in Central Alaska. Human remains and artifacts found at this location are dated to about twelve thousand years ago. Here, researchers found evidence of a more diversified diet that included salmon (Potter et al. 2014).

The term "Amerind" represents about a thousand indigenous languages in the New World (see Supplementary Table 16.3). Thus, it goes without saying that we could not possibly provide a comprehensive discussion of this incredible diversity of Amerind language and culture. Rather, in this section we strive to find examples that will help model the prehistory Amerind languages from a Y-chromosome perspective.

Y-chromosome data for contemporary Amerind populations is limited in terms of resolution. From the published population reports, researchers basically have frequency data for the Q1b-M3 and Q1b-Z780 mutations. Fortunately, ancient DNA from human remains helps to fill the gaps. The available ancient DNA data suggest that the human colonization of the Americas, from Alaska to Patagonia, occurred within perhaps a few thousand years. Given the fact that both regions are separated by a distance of 15 thousand kilometers, the pace of human expansion throughout the Americas occurred fairly rapidly. Evidence for this conclusion comes from the Shuka Kaa remains in Alaska (Kemp et al. 2007) and Sumidouro Cavern remains from Brazil (Moreno-Mayar et al. 2018). Both sets of remains belong to the Q1b-M3 mutation. The dating results from both sites are strikingly similar, about 10 thousand years ago.

This section divides the Amerind data into five regional groups: Pacific Coast, Eastern Woodlands, American Southwest, Central America, Central Andes Mountains, and Amazonia. A synthesis of the data for these groups indicates the following: the pattern and incredible diversity of Amerind languages reflects the diversified and heavily regionalized subsistence strategies of the pre-Columbian Amerind cultures.

5.2. Pacific Coast Indians.

As discussed previously (Section 3), about 15 thousand years ago mammoth hunters crossed over the Bering land bridge from northeastern Asia to Alaska. The mammoth hunters then expanded southwards onto the Great Plains of North America. According to archeological and climatological models, the southward migration onto the Great Plains was facilitated by an ice-free corridor between the Cordilleran and Laurentide ice sheets. Other models, however, favor a second southward migration along the Pacific coastline of the United States and Canada. A recent report (Potter et al. 2018) suggests that although a coastal migration remains plausible, such a model is not supported by the archeological record. Moss and Erlandson in their 1995 paper discuss the terrain along the North American coastline and suggest that mountains, tectonic activity, and the lack of a coastal plain would have hindered a coastal migration. Rather, as suggested by Erlandson, Moss and Des Lauriers (2008) the settlement of the Pacific Coast began with migrations from the North American interior. These migrations then spread westwards alongside rivers that empty into the Pacific Ocean.

The indigenous peoples that settled along the North American Pacific coastline consisted of Amerinds and Athabaskans. They lived in relatively permanent settlements. According to Moss and Erlandson (1995), these settlements exhibit high population density that can be attributed to an abundance of marine resources such as sea mammals and shellfish. The same report divides the indigenous peoples of the North America Pacific coast into three cultural areas: the Alutiiq cultural area, the northwest coast cultural area, and the California cultural area.

The Alutiiq cultural area runs along the coastline of southern Alaska which is the home of the Alutiiq people. They are sometimes referred to as Pacific Eskimos or Pacific Yupik. Their language belongs to the Eskimo-Aleut family.

The northwest coast cultural area begins in southwestern Alaska, where Alaska, the Yukon Territories and British Columbia converge. It terminates in northern California near Fort Bragg. Linguistic diversity within this area is complex. Here, the Eyak-Athabaskan languages have an interesting "leapfrog" distribution. The Tlingit are Eyak-Athabaskan people found at the northernmost part of the cultural area in southwestern Alaska. At the southernmost part, in California and Oregon, are the Pacific Coast Eyak-Athabaskan languages and peoples. In addition to Eyak-Athabaskan languages, linguistic diversity along the northwest coast cultural area includes the Haida language family. Additionally, we find Tsimshian, Wakashan, Salish, and Chimakuan languages and peoples. Finally, we find Ritwan, a sub-branch of the Algic language family. This seems unusual because Algic (or Algonquin) is a major indigenous language family of the eastern United States and southern Canada.

The California cultural area runs south of Fort Bragg and includes the Channel Islands near Santa Barbara. For linguists, the Channel Islands and surrounding region are significant because of its historical association with the Chumash people and languages.

Turning now to the genetic data, contemporary Y-chromosome population data for Pacific coast Indians are limited to the Tlingit and Haida (see Supplementary Tables 16.4 to 16.7). However, we have ancient DNA data for three remains from the California cultural area (Scheib et al. 2018; Supplementary Table 16.8: Reference Nos. 6, 11, and 12). Two remains, one found on St. Miguel Island, and the other near Point Sal, belong to Q1b-M924. Remains from San Nicholas Island belong to Q1b-Y4276. The Q1b-M924 mutation potentially connects the Chumash with Amerinds and an overall model of *in situ* cultural diversification since the onset of the Holocene. The Q1b-Y4276 mutation, on the other hand, may connect the California cultural area with Algic languages. Ritwan provides linguistic support for this argument. Genetic support stems from Grugni et al. (2019). This study identifies Q1b-Y4276 as a potential marker for Algic peoples of the northeastern United States. Alternatively, Q1b-Y4276 may link the California cultural area with Eyak-Athabaskan. A downstream variant of Q1b-Y4276, the Q1b-B34 mutation, was found in ancient remains at an Athabaskan cemetery near Kenai, Alaska (Scheib et al. 2018; Supplementary Table 16.8: Reference No 7).

5.3. Eastern Woodlands.

We define the Eastern Woodlands as the United States east of the Mississippi River, as well as southern Ontario and Quebec. From an archaeological perspective, this region was inhabited around the beginning of the Holocene (Snow 2013: 354). Unlike the indigenous peoples of the Pacific coast of North America, the indigenous peoples of the Eastern Woodlands supplemented their subsistence strategy with farming. Farming within the Eastern Woodlands may have fueled an expansion of the Algic, Iroquoian, Siouan-Catawban, and Muskogean language families (Bellwood 2005: 174-179). Crop cultivated in this region includes maize, beans, squash, sunflower, tobacco and goosefoot.

At this point we return to the Algic language family which is often called "Algonquin." Campbell (1997:156) places the putative homeland of Algic languages around the Great Lakes but notes that some have placed it further west. Algic languages are distributed over a vast area. As previously mentioned, this language family is found in the eastern United States and northern California. Moreover, we should mention that Algic languages are distributed across much of southern Canada. Finally, some of the Algic-speaking peoples, such as the Cheyenne and Arapahoe, inhabited the Great Plains of the United States.

A recent study (Grugni et al. 2019) reports that the Q1b1a1a2-Y4276 mutation is distributed from Siberia to South America. The same study suggests this mutation is a potentially useful marker for Algic languages. However, the assertion is not supported by published data. Rather, published studies of contemporary indigenous populations of North America report frequency data for Q1b-M3, Q1b-Z780 and C2b-P39 (see Supplementary Tables 16.4 to 16.7).

Another large linguistic family of the Eastern Woodlands is Iroquoian. These languages are found in the vicinity of the Great Lakes and extend southwards along the Appalachian Mountains to Georgia. Bolnick et al. (2006) suggest, based on their analysis of the genetic data, that the putative homeland of Iroquois-speaking peoples is the southeastern United States. However, linguistic and archaeological perspectives place the Iroquoian homeland in the Appalachian uplands, which encompasses a vast area from Pennsylvania to Georgia (Snow 2013: 358).

Snow (2013: 359-360) describes indigenous peoples that inhabited the Mississippi River valley and the lower Ohio River. These peoples include those that speak languages belonging to the Siouan-Catawban and Muskogean language families. According to Snow, around the year 1000 AD many of the Siouan-Catawban peoples, such as the Mandan, practiced what appears to be intensive agriculture. However, a drought around 1450 AD pushed some of the Siouan-Catawban people like the Lakota onto the Great Plains where they abandoned farming altogether.

5.4. The American Southwest.

We define this area as southern California, Nevada, Arizona, New Mexico and southwestern Colorado. Linguistic diversity in this area includes the Uto-Aztecan, Kiowa-Tanoan, Eyak-Athabaskan language families and the Zuni language isolate. Haplogroup Q-M242 data are available for the Jemez, Tohono O'odham (Papago), Akimel O'odham (Pima), Navajo, and Apache peoples (see Supplementary Tables 16.4 and 16.5). C2b-P39 has also been reported for the Navajos and Apache (see Supplementary Table 16.7). The Jemez language belongs to the Kiowa-Tanoan language family. The Tohono O'odham, spoken by the Pima and Papago, are among the Uto-Aztecan-speaking peoples. Navajo and Apache are classified as Eyak-Athabaskan languages.

The Uto-Aztecan family has a vast geographical distribution, from Oregon in the United States to Panama (Campbell 1997: 133). In order to understand the evolution of Uto-Aztecan, we must discuss language variation in Central America, which is detailed below. At this point we note that the Uto-Aztecan family consists of sixty-one languages. The family has two main divisions: Northern Uto-Aztecan and Southern Uto-Aztecan. Northern Uto-Aztecan consists of thirteen languages found in the United States. Examples include Hopi, Comanche, Shoshoni and Paiute. The Southern Uto-Aztecan branch consists of forty-eight languages. Forty-seven of these languages, such as Nahuatl, the language of the Aztecs, are found in Central America. Tohono O'odham is the only Southern Uto-Aztecan language found in North America.

Agriculture may have played a role in the expansion of Uto-Aztecan languages. Additional details will follow below in the discussion of Central America. In the meantime, it is necessary to discuss Numic languages, a sub-branch of Northern Uto-Aztecan. Its speakers include the Comanche, Paiute, Mono, and Shoshoni peoples. According to the archaeological record, it appears as though they abandoned farming about a thousand year ago and adopted foraging as their subsistence strategy (LeBlanc 2013: 373). This reversion to foraging helps to illustrate the correlation between linguistic variation and subsistence variation, a point that we make in this report.

5.5. Central America.

For the purposes of this discussion, the border of the United States and Mexico defines the northern boundary of Central America. The border of Panama and Colombia defines the southern boundary. In Central America, Y-chromosome population data is available for the following language families: Chibchan, Chocoan, Mayan, Mixe-Zoquean, Otomanguean, Tarascan, and Uto-Aztecan (see Supplementary Tables 16.9 to 16.11.

All the Uto-Aztecan languages of Central America are classified within the Southern Uto-Aztecan branch. The Otomanguean family consists of 178 different languages found in Mexico. Examples of Otomanguean languages include Mixtec, Zapotec and Otomi. The Mayan family consists of thirty-one languages found in Mexico and Guatemala. Mayan is considered a linguistic relic of the Mayan civilization. Seventeen languages are classified within the Mixe-Zoquean language family of Mexico. Campbell and Kaufman (1976) suggest this language family is a linguistic relic of the Olmec civilization. The Chibchan family consists of twenty languages which are found in Costa Rica, Panama, Honduras, Nicaragua, and Columbia.

Bellwood (2005: 237-239) provides a short discussion of the co-evolution of farming and language in Central America. He suggests that early maize cultivation fueled expansion of the Mayan, Otomanguean, and Mixe-Zoquean language families. Bellwood (2005: 240-244) then discusses the Uto-Aztecan language family. He takes the position that this expansion follows the early farming dispersal hypothesis. This opinion was shaped by collaboration with the anthropologist Jane Hill. In a paper published in 2001 she suggests that Uto-Aztecan speakers were among the early maize farmers of Mexico. Around six thousand years ago as the result of population pressure they began to expand northwards. Between three and four thousand years ago they migrated into the American Southwest and continued to cultivate maize and other crops. Hill supports her model mostly with linguistic reconstructions. A study from 2010 (Kemp et al) supports her model with Y-chromosome data, the distribution of Q1b-M3 and Q1b-Z780 mutations as well as short tandem repeats (STR).

An alternative interpretation of the data analyzed by Kemp et al (2010) would posit the absence of a unique genetic signature for Uto-Aztecans or any of the other Central American language families. The best resolution markers we have include Q1b-Y12421, which represents the majority of Q-M3 variation among Panamanians; Q1b-M924, which represents most of the Q-M3 variation in Mexico; Q1b-Z5906, which is distributed from Mexico to Argentina with a peak frequency in Peru; and Q1b-Z5908, which is distributed from Mexico to Argentina with a peak frequency in Peru (Grugni et al. 2019). At best, these recently reported markers merely suggest an increase in Central and South America population beginning about five thousand years ago. However, we cannot build farming-language expansion models with the currently available Y-chromosome data.

The position taken by Hill (2001) is controversial because many researchers (e.g. Campbell 1997: 150) place the putative homeland of Uto-Aztecan languages somewhere in the southwestern United States or northern Mexico. Hill, on the other hand, places the origins of Proto-Uto-Aztecan much further south, somewhere in south-central Mexico where maize was first cultivated.

The Uto-Aztecan language-farming expansion, as posited by Hill (2001), was contested in a 2009 paper. Merrill et al. asserted that phonological reconstructions for flora and fauna place the putative homeland in Nevada and not in southern Mexico. Based on climatological data, the researchers further assert that a drought led to a bifurcation of Proto-Uto-Aztecan into the Northern and Southern Uto-Aztecan branches about nine thousand years ago. Southern Uto-Aztecan then expanded southwards from Nevada into Mexico. They then suggest, based on their analysis of climatological and archeological data, that a Southern Uto-Aztecan group back-migrated from Mexico into the southwestern United States occurred about six thousand years ago. According to the report, this back-migration brought domesticated maize from Mexico into the region. Finally, Merrill et al. (2009) suggests that this expansion of maize and language was fueled by climate change rather than population pressure.

Extending the discussion above, maize became an important food resource among many of the Native American cultures. It was the only grain-like resource of the Western Hemisphere that could be stored for a long period of time. However, the road to a food staple was a long and complicated process that required considerable genetic modification of teosinte, the wild plant from which modern domesticated maize evolved. A recent study (Kistler et al. 2018) examined the domestication of maize by using a synthesis of genetic, archaeological and botanical data. The study reports that domestication began roughly nine thousand years ago in south-central Mexico. However, according to the study even 5.3 thousand years ago the Mexican variant of maize had not evolved into a food staple. Thus the proposed timing of northward co-expansion of Southern Uto-Aztecan and maize, as suggested by Merrill et al. (2009), six thousand years ago, is problematic. Perhaps the expansion of Uto-Aztecan and maize cultivation into the southwestern United States is not related.

Smalley and Blake (2003) provide a useful discussion of maize origins from botanical and anthropological perspectives. As previously mentioned, modern domesticated maize evolved from the wild teosinte plant. The report notes that teosinte cobs are much smaller than modern maize. Moreover, the kernels are barely edible. The study even describes teosinte kernels as "starvation" food that is otherwise "utterly useless." As such, this poses an interesting question: why would anyone waste so much time and energy to cultivate such a plant? According to Smalley and Blake (2003), the answer is alcohol. The teosinte stalks are sweet and initially people chewed them. Eventually someone discovered that when pressed the stalks yield syrup that can be used for corn wine. As such, people initially cultivated maize as a recreational product rather than for food. According to Smalley and Blake (2003), during the recreational phase of maize domestication, farmers planted seeds that they had gathered from the larger maize stalks with the idea of obtaining a larger yield of syrup with the next harvest. This selection of seeds from larger stalks eventually produced the large cobs that are characteristic of modern domesticated maize. At this point people started to dry maize kernels and maize became a food staple throughout the Native Americas. Nevertheless, some continued to produce alcohol from maize by using the kernels for making beer (e.g. *chicha*).

5.6. Central Andes.

Heggarty and Beresford-Jones (2010) define the Central Andes region as the central Peruvian highlands and the western Pacific coastline of Peru. Researchers have proposed that the Quechuan and Aymaran language families evolved in this region (Bellwood 2005: 235; Heggarty and Beresford-Jones 2010). In terms of number of speakers, Quechuan represents the largest of Native American languages. According to *Ethnologue* (2019) around 7.8 million people speak one of the forty-four Quechuan languages. Aymaran represents a smaller language family with three languages and around 1.7 million speakers. Mainstream linguistic opinion (e.g. Campbell 1997: 188) does not propose a genealogical relationship for both language families despite a large shared vocabulary and many structural similarities.

Very solid archeological and genetic evidence place *Homo sapiens* in South America by at least 10 thousand years ago (Roosevelt et al. 1996; Moreno-Mayar et al. 2018; Capriles 2019). From a Y-chromosome perspective the genetic relics of this migration are the Q1b-M3 and Q1b-Z780 mutations (see Supplementary Tables 16.12 to 16.14). Downstream from Q1b-M3 marker, several mutations point to substantial population expansion with the Central Andes in the last five thousand years (Jota et al. 2016; Grugni et al. 2019). Agriculture probably fueled the expansion. The transition to agriculture in the Central Andes included the domestication of plants and animals. Domesticated plants include potatoes, sweet potatoes, quinoa, and maize. Alpacas, vicuñas, alpacas and llamas, which are classified as camelids, represent the domesticated animals.

In order to understand the agricultural transition in the Central Andes, a discussion of geography is necessary. One finds a very steep rise in elevation. The western coast of Peru lies at sea level. In the central highlands, the elevation can reach six thousand meters. During the Pre-Ceramic phase, about 11 thousand to about 4.5 thousand years ago, human activity was concentrated along the coastline (Heggarty and Beresford-Jones 2010). The abundance of marine resources appears to have drawn people to this area. However, archeological remains, radio-carbon dating, and stable oxygen isotope data (Haas et al. 2017) suggest that coastal hunter-gathers made seasonal treks into the highlands. They hunted this region in order to harvest wild camelids such as alpacas. Then around seven thousand years ago humans occupied the highlands on a permanent basis.

Over time the coastal and highland people of the central Andes became increasingly dependent on agriculture and less dependent on foraging. The agricultural transition in the highlands included the domestication of the camelids that they once hunted. This resource provided meat as well as fleece for clothing. Additionally, highlanders utilized these animals as beasts of burden (for more details, see Mengoni-Gonalons and Yaco-Baccio 2006). The road to agriculture in the Central Andes also included the cultivation of crops. Potatoes became an important cultivated food resource, both in the highlands and along the coast. Since modern potatoes consists of a large number of hybrids and variants, identifying how and exactly when this food resource arrived in the Central Andes is problematic. What we know stems from studies that analyze genetic data (Spooner et al. 2005; Hardigan et al. 2017). They suggest that the potato might have been domesticated in southern Peru about eight to ten thousand years ago.

For coastal people the sweet potato was in addition to potatoes an important crop. Like the potato, identifying the how, where and when of modern sweet potatoes is problematic because the large number of hybrids and variants. In terms of location, it appears that this crop could have evolved independently in the Caribbean, Central America and northwestern South America (Roullier et al. 2013).

Maize became an important food resource in the Central Andes around three thousand years ago. This crop was cultivated both in the highlands and lowlands. Heggarty and Beresford-Jones (2010) suggest that the cultivation of this crop signals an intensification of agriculture within the region. The intensification of agriculture eventually produced population pressure and soil degradation along the coast. As a result, coastal settlements were abandoned and the highlands became the focal point of human activity (Bellwood 2005: 163-164).

Finally we should mention quinoa, a type of chenopod that is often confused as grain rather than a source of edible seeds. Quinoa became an important food resource in the highlands of the Central Andes where it was domesticated roughly three thousand years (Bruno 2006).

The linguistic prehistory of the Central Andres was shaped by several different cultural transitional periods that arose between the adoption of maize (around three thousand years ago) and the arrival of the Spanish in 1532. Heggarty and Beresford-Jones (2013: 405) describe Aymaran languages as a linguistic relic of the Chavin culture and the Early Horizon period, roughly 900 BC to 100 AD. Both researchers describe Quechuan as a linguistic relic of the Wari civilization and the Middle Horizon Period, roughly 550 AD to 1000 AD.

In their 2010 paper, Heggarty and Beresford-Jones consider Bellwood's early farming dispersal hypothesis (2005: 1-11). They acknowledge that this model helps to decipher linguistic evolution in the "Old World." However, according to the researchers the model is problematic in the Central Andes. Instead of a co-expansion of language and early agriculture, as predicted by the early farming dispersal model, linguistic variation in the Central Andes conforms to a model of *in situ* co-evolution of language and farming.

At this point we note that *in situ* co-evolution of language and farming also occurred in the Old World. Japonic and Koreanic are two very good examples (see Chapter 15: Sections 14 and 15. However, the transition to intensive agriculture leveled linguistic diversity in both regions. In the Central Andes, on the other hand, the same leveling of linguistic diversity seems not to have occurred. Rather, Heggarty and Beresford-Jones (2010) suggest diglossia within the region. When the Spanish arrived, Quechua was the high variety and Aymara the low. Perhaps this diglossia reflects the absence of intensive agriculture for a sufficient period of time. Extending this argument further, the Spanish may have interrupted what would have ultimately been a natural leveling of language diversity that is characteristic of intensive agriculture.

The idea that agriculture follows a gradient of intensification was explored by Stevens and Fuller in their 2017 paper. They suggest that the transition to agriculture only occurs when a population obtains fifty percent of its calories from domesticated plants and animals. According to the report, the road to agriculture can have a lengthy pre-agricultural phase. During this phase, hunter-gathers often cultivate crops on a smaller scale. However, this is not agriculture. Rather, as suggested by the study, the transition to agriculture essentially marks a point-of-no-return. Agriculture vastly improves productive success, and this comes with a price. At this point foraging is no longer an option because you must feed many more people. Furthermore, habitat for wild animals and plants are now utilized as farmland. Finally, intensive agricultural eventually creates social institutions that undermine linguistic diversity.

5.7. Amazonia.

Amazonia is usually associated with the world's largest rainforest. For the purpose of this discussion, the geography of this region is defined by the Orinoco and Amazon Rivers and the vast number of tributary rivers that flow into them (see Figure 16.2). The archaeological record suggests that Amazonia has been inhabited for at least 10 thousand years (e.g. Roosevelt et al. 1996; Capriles 2019). This closely follows dating estimates acquired from ancient DNA data retrieved from the Sumidouro Cavern in Brazil (Moreno-Mayar et al. 2018), which provide the most robust time estimates for the human settlement of South America.

Amazonia is linguistically complex. Major language families of the region include the Carib, Tupi, Panoan, Jean, Tucanoan and Arawak language families. In this discussion of Amazonia we will focus on Arawak. The prehistory of this language family is strikingly similar to that of Austronesian. As such, Arawak seems to provide good example of a New World language family that conforms to the *early farming dispersal hypothesis* as postulated by Bellwood (2005: 1-11).

Before discussing the prehistory of Arawak we focus briefly on the modern distribution of this language family. Data for this discussion comes from *Ethnologue* (2019). The organization utilizes the alternate name for Arawak, which is Maipurean. *Ethnologue* lists fifty-six different Maipurean languages and roughly three quarters of a million speakers. Maipurean has two main divisions, a northern branch and a southern branch. Southern languages are found in Peru, Bolivia and Brazil. The northern branch is found in Brazil, Suriname, Guyana, Columbia, Venezuela, Puerto Rico and Honduras.

In 1492 Arawak was a linguistic heavyweight within Amazonia and in the Caribbean. Arawak languages thrived and survived because the Arawak people had mastered the art of tropical agriculture along major river systems. Part of their success stems from the construction of raised field agriculture. Amazonian rivers tend to flood regularly. By constructing fields above the floodplain, they greatly increased the efficiency of agriculture by ensuring adequate drainage and improving the fertility of otherwise poor growing soil (Whitney et al. 2014). Another factor is crop selection which includes sweet potatoes and maize. However, the most important crop was manioc. Also known as cassava, manioc is a root vegetable that can be cultivated in the poor-quality soil of tropical climates. Another advantage is that bitter manioc can be made into a flour and stored.

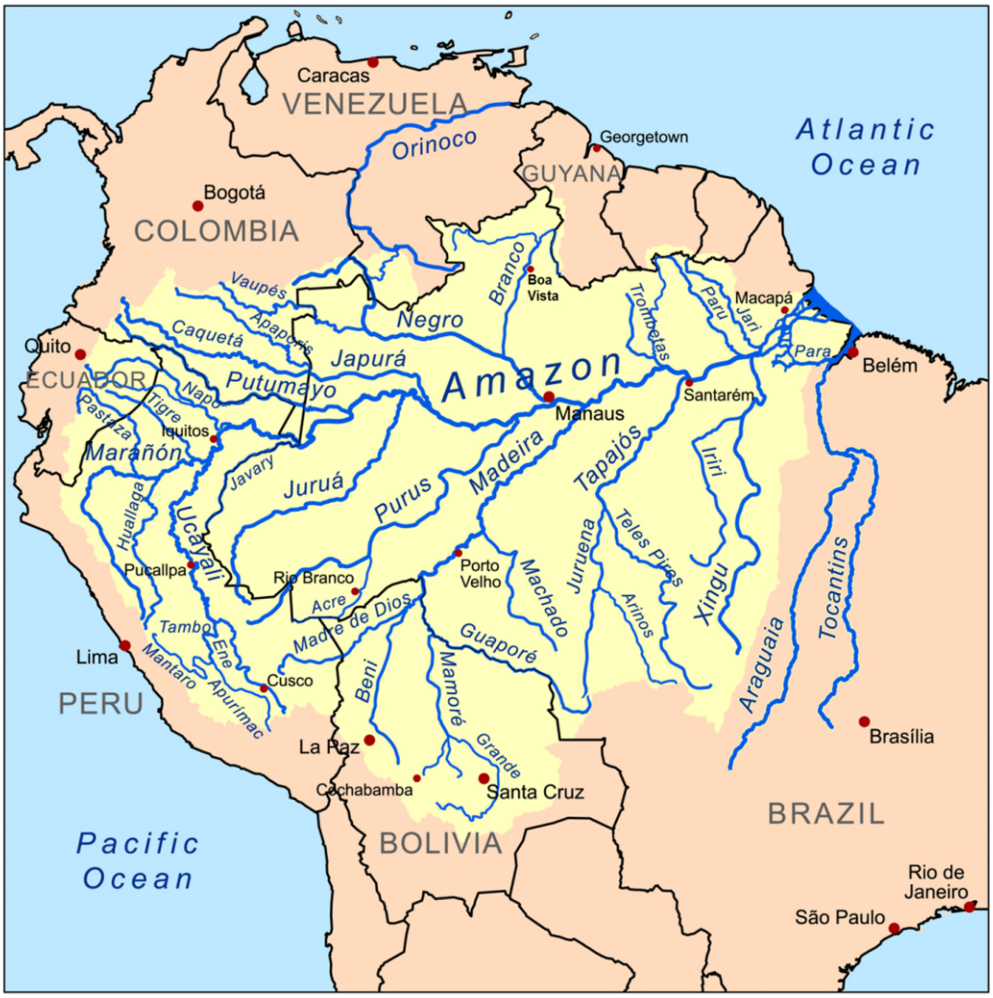


Figure 16.2. Amazon River Basin.
File licensed under Creative Commons Attribution-Share Alike 3.0 Unported license.Link to File

At this point we examine a potential co-expansion of Arawak and raised field agriculture from southwestern Bolivia about 2.5 thousand years ago. In doing so we depart from mainstream linguistic opinion that places Arawak origins further north (e.g. Aikenvald 1999). Our view of Arawak origins is supported by the archeological and botanical data (Whitney et al. (2014) that places the origins of raised field agriculture in the Llanos de Moxos region of Bolivia about 2.5 thousand years ago. This dating estimate for raised field agriculture corresponds to a massive expansion of Arawak settlements alongside the numerous rivers of Amazonia (Horborg 2005; Heckenberger 2013). Furthermore, the South American variant of domesticated maize traces its origins within or near this region (Kistler et al. 2018). Finally, the region is a potential domestication center for manioc (Olsen and Schaal 1999).

Similar to the "Austronesian advantage" that evolved in Island Southeast Asia, the "Arawak advantage" evolved in South America. Researchers have asserted that foraging cannot sustain human population in tropical rainforests such as Amazonia. Rather, people need to supplement their diet with agriculture (Bailey et al. 1989). Such a position seems contrary to the archaeological data (e.g. Roosevelt 1996) which, in fact, record pre-agriculture occupation of Amazonia. Nevertheless, one still finds compelling arguments for the idea that foraging is not capable of sustaining high population density within the tropical rainforests. Thus, it seems significant that Horborg (2005) describes the Pre-Columbian Arawak settlements or villages as chiefdoms with high population density. The so-called "Arawak Advantage" points to cultural adaptations, such as raised field agriculture and plant domestication, that drove greater reproductive success, which drove the expansion of Arawakan languages.

When Columbus landed in the New World, among the first indigenous peoples he encountered were the Taínos, speakers of the Taíno language, which belongs to the Arawakan language family. Taínos were descendants of a second Arawak expansion that began roughly 2.5 thousand years ago from the northern coast of South America. Initially, the Taínos settled the Lesser Antilles Islands. Later, they expanded into Hispaniola, Puerto Rico, the Bahamas, Jamaica and Cuba. This expansion carried many of cultural features of Arawak cultures found on the South American mainland: the intensive cultivation of manioc; the dominance of large trade networks; villages centered around a plaza; pottery; social organization; and high population density (see Wilson 2007: 59-136; Keegan 2013: 376-383 for a more detailed discussion).

Dixon and Aikenvald (1999: 7) estimate that between 2 and 5 million people lived in Amazonia prior to the arrival of Europeans. According to the report, since 1492 European diseases and population displacement have significantly altered the linguistic landscape within this region. It is difficult to reconstruct what was once there. The researchers further note that the surviving indigenous languages of the region remain understudied in academia. Similar to the paucity of linguistic data, one finds a very limited amount of Y-chromosome data. We mostly have frequency results for Q1b-M3 and Q1b-Z780, high resolution markers that are not particularly informative (see Supplementary Tables 16.12 to 16.14). Furthermore, we note that Native American Y-chromosome lineages have disappeared among Caribbean populations (see, for example, Marcheco-Teruel et al. 2014).

With respect to modeling the prehistory of Arawakan languages, the lack of linguistic and genetic data can be mitigated by examining the prehistory of Austronesian languages (see Chapter 15: Sections 6 to 10 for more details. The similar prehistory of both language families is striking. Austronesians and Arawaks excelled at navigation. As a result, both groups dominated regional trade alliances. Austronesians and Arawaks excelled at tropical agriculture. Both groups cultivated tubers that grow in the tropic environment: taro in the case of Austronesians, and manioc in the case of Arawaks. Both groups farmed were nobody could farm. An evolutionary adaptation allowed Austronesians to farm in malaria infested lowland coastal regions. Arawaks perfected riverine agriculture by constructing raised fields above the flood plain. Their success at agriculture fueled rapid population growth, which fueled a rapid co-expansion of people and language.

Section 6. Athabaskans.

The reader should note that the terms Na-Dené and Eyak-Athabaskan are essentially synonymous. Greenberg (1987: 321-330), for example" describes the so-called "Na-Dené problem." However, other linguists such as Campbell (1997: 110-155) and *Ethnologue* (2019) use Eyak-Athabaskan. Since Eyak-Athabaskan seems to reflect consensus among the linguists, we utilize it also. According to Ethnologue (2019), the Eyak-Athabaskan language family consists of forty-four different languages. These languages have a leap-frog distribution over a vast geographic range (see Figure 16.3 below).

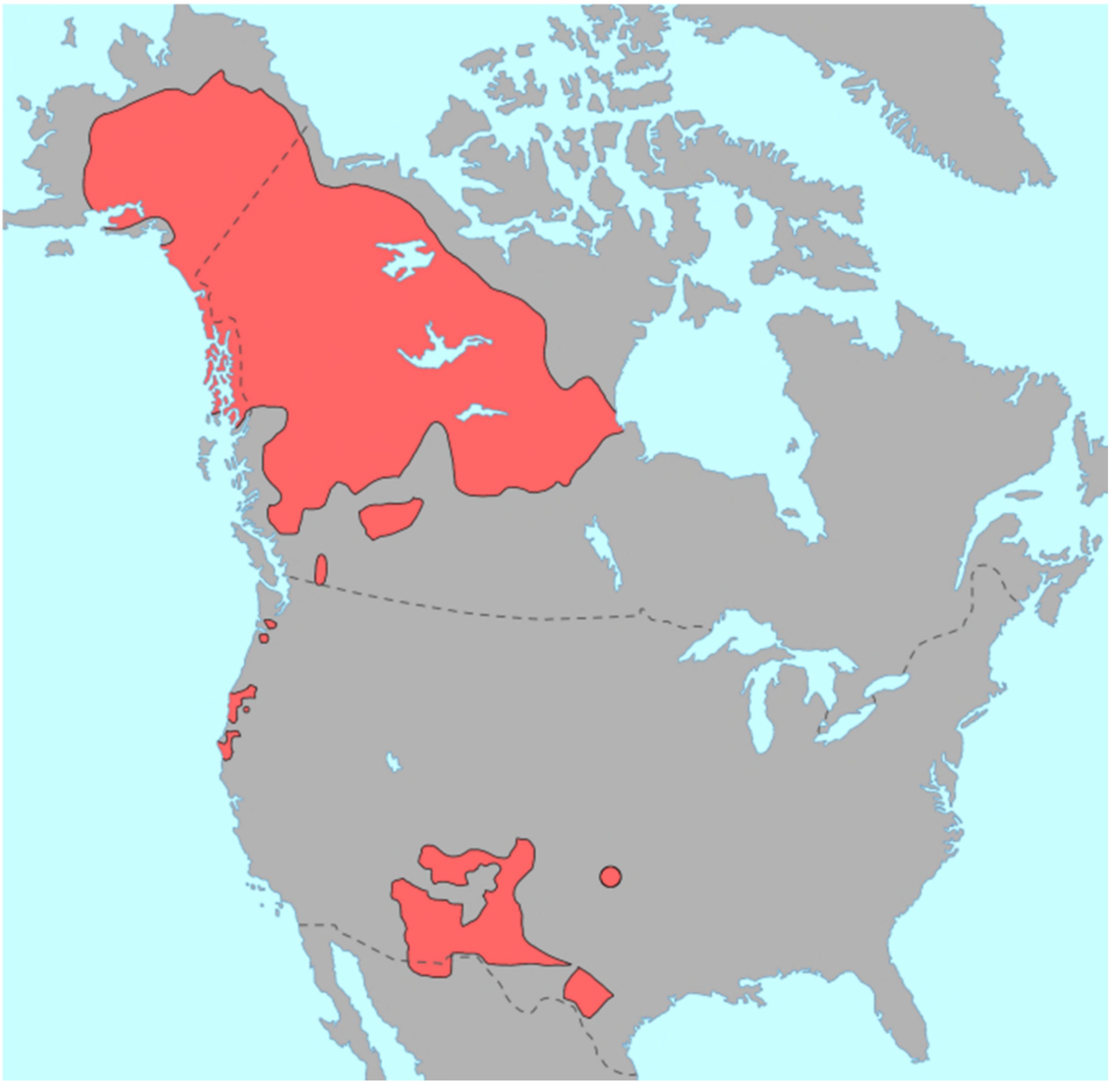


Figure 16.3. Distribution of Athabaskan Languages.
File licensed under the Creative Commons Attribution 2.0 Generic license. Link to File

Ethnologue (2019) divides the Eyak-Athabaskan language family into three main divisions: Eyak, Athabaskan and Tlingit. Eyak is a single language that is now extinct. It evolved near the mouth of the Copper River in southern Alaska. Tlingit is a single language branch from the coastal region of southeastern Alaska. The Athabaskan branch has three sub-branches: Apachean, Northern Eyak-Athabaskan, and Pacific Coast Eyak-Athabaskan. Apachean languages are found in the desert of the southwestern United States. This sub-branch consists of the Navajo and Apache. Turning now to the Northern sub-branch, here we find twenty-seven different languages distributed throughout Alaska and Canada. Finally, the Pacific Coast sub-branch of Athabaskan consists of languages along the Oregon and Californian coast in the United States.

From a linguistic perspective, the Eyak-Athabaskan language family potentially arose within the interior of North American, where Alaska, British Columbia and the Yukon converge on the map. Despite close geographic proximity in Alaska and Canada, Eskimo influence in Eyak-Athabaskan is negligible. Haida and Eyak-Athabaskan, on the hand, borrowed from each other. However, the data fail to support a genealogical relationship for both language families (for additional details, see Campbell 1997: 110-115).

From an archeological perspective, Gillispie (2018) suggests that the Athabaskan cultural tradition arose in interior Alaska around 1700 years ago. According to the researcher, the appearance of the cultural tradition corresponds to a technological innovation in the region, the bow and arrow. Interestingly, Gillispie (2018) suggests Eyak, Tlingit and Haida societies evolved before the Athabaskans at around 2500 years ago. This estimate corresponds to climate change that stabilized coastlines, as well as cooler weather and greater precipitation. Thus, from an archaeological perspective, Eyak-Athabaskan may have origins along the southern Pacific coast of Alaska rather than in the Alaskan and Canadian interior.

Matson and Magne (2013) time the Athabaskan expansion into interior Alaska and British Columbia with the Mount Churchill volcano eruption around 300 AD. Both researchers suggest that a second more powerful eruption around 800 AD. This eruption drove Athabaskan peoples either into the northwestern Pacific coast of the United States or into the American Southwest. Linguistic support for this expansion model comes from the Dakelh people of British Columbia. They speak Carrier, an Athabaskan language that is closely related to Apachean languages.

Very little contemporary Y-chromosome data exists for Eyak-Athabaskan populations (see Supplementary Tables 16.4 to 16.7). These data stem from a 2012 study (Schurr et al.) that waded into the long-standing linguistic debate about the classification of Haida and Eyak-Athabaskan. The study was not able to provide genetic evidence of a common ancestral population for these populations. Nevertheless this study along with others underscores the unexpected presence of the rare C2b-P39 mutations among Eyak-Athabaskan populations which include the Tlingit of southeastern Alaskan coast; the Tanana of interior Alaska; the Dogrib and Gwich'in of Canada; and the Navajo and Apache of the southwestern United States (see Supplementary Table 16.7).

It should be emphasized that C2b-P39 has also been detected in Algic, Eskimo-Aleut, Iroquoian, Muskogean, and Siouian-Catawban speaking populations (see Supplementary Table 16.7 for additional details). Given the contemporary distribution of C2b-P39, this mutation represents a potential founder lineage for North America. In other words, some of the Paleo-Siberians had the mutation when they crossed over the Bering land bridge roughly 15 thousand years ago. Such a position was recently taken by Wei et al. in their 2018 report that analyzed Asian C-M130 lineages that are closely related to C2b-P39. However, an alternate scenario would suggest that C2b-P39 represents more recent geneflow between Alaska and the Kamchatka Peninsula as suggested by Pinotti et al. (2019). Such a scenario is supported by the presence of the C2b-FGC28881.2 mutation found in Koryaks (Wei et al. 2017b). The C2b-FGC28881.2 mutation is the closely related C2b-P39 sister clade mutation.

C2b-P39 data from contemporary populations, along with Q-M242 data from contemporary and ancient DNA studies, potentially support prehistoric contact between the indigenous peoples of northeastern Asia and the Athabaskans of Alaska. Prehistoric contact is a salient point for linguists because some researchers have suggested a close linguistic relationship between the Yeniseian language family of south-central Siberia and the Eyak-Athabaskan language family. This close relationship was proposed by Merritt Ruhlen in 1998 based on thirty-six cognate sets which include basic vocabulary. Nevertheless, this conclusion has proven controversial among the linguists. Campbell (2011) asserts a lack of linguistic evidence. The geographic distance between Yeniseian and Athabaskan speaking populations is also a problem.

We now turn to the Q1b-L330 mutation. As illustrated by Supplementary Figure 16.1: Part B, the Q1b-L54 mutation splits into Q1b-M930, Q1b-Z780 and Q1b-L330. The Q1b-L330 mutation is not found in Native Americans. Rather, it is confined to Siberia where it is the predominate mutation of the Ket people (Flegontov et al. 2016b; Wei et al. 2018). The Kets are the sole source of genetic data for Yeniseian languages. As such, the available Y-chromosome data fail to support the controversial Yeniseian-Athabaskan hypothesis.

Finally, we turn to the linguist Joseph Greenberg. He suggested (1987: 323) that Amerind represent an initial migration into the Americas, and that the Eyak-Athabaskan represents a second migration. From archaeological and linguistic perspectives, a second Eyak-Athabaskan migration into the Americas seems unlikely. However, we cannot reject this proposal with genetic evidence due to a lack of data (see Section 8 for more details).

Section 7. Eskimo-Aleut.

According to *Ethnologue* (2019) the Eskimo-Aleut language family consists of eleven languages. This language family has two main divisions, Aleut, a single language branch, and the Eskimo branch with the remaining ten languages. The Aleut language is found on the Aleutian Islands. Eskimo has two sub-branches, Inuit-Inupiaq with five languages and Yupik with five languages. The geographic distribution of Inuit-Inupiaq languages follows the Alaskan coastline north of Unalakleet along the Bering Sea and Artic Ocean. They further extend along the Arctic Ocean coastline in Canada into Hudson Bay. From Hudson Bay, Inuit-Inupiaq extends into Greenland. The Yupik sub-branch is found on both sides of Bering Sea. Two languages, Naukan and Sirenik, are spoken on the Chukotka Peninsula in Russia. Three Yupik languages are spoken the United States. The Central Yupik language is found along the Bering Sea Coastline of western Alaska. Pacific Yupik is spoken along the Pacific coastline of southern Alaska. St. Lawrence Yupik is found on St. Lawrence Island in the Bering Sea.

The reader is directed to the following link that provided a map of the indigenous languages of Alaska: http://www.alaskool.org/language/languagemap/index.html. In his reference guide to Native American languages, Lyle Campbell (1997: 353) also provides a map that illustrates the distribution of the Eskimo-Aleut language family. In the same reference guide (1997: 109), Campbell places the geographic point of origin of this language family in southwestern Alaska near Bristol Bay and the Cook Inlet. In the discussion he rejects a close linguistic relationship between Eskimo-Aleut and the Uralic family of Northern Eurasia. Similarly, a common ancestral language for Eskimo-Aleut and the Chukotko-Kamchatkan family is also problematic.

Turning now to the archaeological perspective, the where and when of Eskimo-Aleut origins is confusing. A discussion of this language family begins with the Paleo-Eskimos, the so-called Dorset cultural tradition and their expansion across the Arctic Ocean of North America. The Paleo-Eskimo expansion may represent a secondary expansion of the Arctic Small Tool tradition. Tremayne (2015) places the origins of the Arctic Small Tool tradition in circumpolar region of northeastern Asia at around 5 thousand years ago. By around 3 thousand year ago, the tradition had arrived in Alaska where it spread southwards to Kodiak Island and eastwards along the Arctic Ocean (Dumond 2005; Friesen 2013: 346-349).

As previously stated in Chapter 14: Section 14.9, about fifty percent of Siberian Yupik have the N-M231 haplogroup whereas the mutation is absent among North American Eskimos. Perhaps the arrival of reindeer herders in northeastern Siberia and population pressure drove some of the Paleo-Eskimos across the Bering Sea into Alaska and beyond. This would assume that Eskimo-Aleut peoples predate the arrival of reindeer domestication in northeastern Siberia. After crossing the Bering, the archaeological record (Gillispie 2018: 30) suggests that the Paleo-Eskimos of North America were highly mobile foragers. They alternated their subsistence strategy between inland and coastal resources. In the winter they settled along the coast to hunt seals. When the weather became warmer, they moved inland to intercept migrating herds of caribou and muskoxen.

The Neo-Eskimo/Thule tradition eventually replaced the Paleo-Eskimo tradition. Fortescue (2013: 341) and Gillispie (2018: 23) suggest that this occurred by around one thousand years ago. This transition involved significant cultural changes. The Thule became successful whale hunters. As a result of this food resource, they built permanent settlements and focused on marine resources. This resource also helped to increase population density. Permanent settlements and greater population density eventually brought more complex social structures, and with that, trade alliances and warfare (Friesen 2013: 349-351).

According to Fortescue (2013: 340) linguistic relics of the Paleo-Eskimos have disappeared. This suggests that the Thule tradition involved a population expansion and potential assimilation of the Paleo-Eskimos. The previous discussion of the geographic origins of the Eskimo-Aleut, and the distribution of the languages of this family, suggest that the Thule expansion began in southwestern Alaska. Eskimo-Aleut languages then radiated in several directions: westward into the Aleutian Islands and northeastern Asia; eastward along the southern Alaska coast; northwards along the eastern Bering Sea coastline of Alaska; and finally, along the Arctic Sea coastline of North America. However, the available genetic data paint a different picture. In terms of time depth, Eskimo-Aleut languages may extend much further back in time to the Paleo-Eskimos. Additionally, the genetic evidence may place the origins of Eskimo-Aleut in northeastern Asia (see Section 8 below for additional details).

Section 8. Bi-Directional Flow of Language, Genes and Culture across the Bering Sea.

8.1. Overview.

Due to an absence of published genetic data for contemporary Native Alaskan populations, it is not possible to determine the extent of genetic admixture between Eyak-Athabaskan and Aleut-Eskimo speaking peoples. Nevertheless, the available archeological and linguistic data suggest that Eyak-Athabaskan and Eskimo-Aleut speaking peoples maintained considerable cultural distance. The Eyak-Athabaskans focused on the resources of interior Alaska. Eskimo-Aleuts, on the other hand, exploited marine resources along the coast (see Gillispie 2018 for a more detailed discussion). Additionally, the available data suggest extensive bi-directional flow of language, genes and culture across the Bering Sea. The indigenous Paleo-Siberian peoples of northeastern Asia may have shaped the evolution of the Eyak-Athabaskan and Eskimo-Aleut language families.

8.2. Linguistic Evidence of Bidirectional Flow.

As discussed previously in Section 6, some researchers suggest that a common ancestral language may link Eyak-Athabaskan and Yeniseian. However, this proposal is also controversial among the linguists. Far less controversial is the idea that the Eskimo-Aleut family was shaped by the indigenous peoples of North America and northeastern Asia. Linguists agree that Eskimo-Aleut languages are spoken on both sides of the Bering Sea. Moreover, linguistic evidence may suggest that Eskimo-Aleut may have been spoken on the Kamchatka Peninsula by the coastal Chukchi people (Fortescue 2004). It should be noted that they now speak a Chukotko-Kamchatkan language.

8.3. Anthropological Evidence of Bidirectional Flow.

From an anthropological perspective, whale hunting appears to have mediated long-term cultural exchange along the northern Pacific Rim. Savelle and Kishigami (2013), in their discussion of prehistoric subsistence whaling, draw a distinction between opportunistic and active whaling. According to the researchers, Jomon archeological sites in Japan provide evidence of opportunistic whaling at around nine thousand years ago. Intensive or active whaling then began about five thousand years ago on the Noto Peninsula in Japan. Active whaling eventually spread northwards through the Kurile Islands. By around three thousand years ago, active whaling reached the Kamchatka Peninsula and Chukotka (see Figure 16.4). Perhaps as early as fifteen hundred year ago, whaling reached Alaska. Finally, by around eight hundred years ago, whaling had advanced across northern Canada.

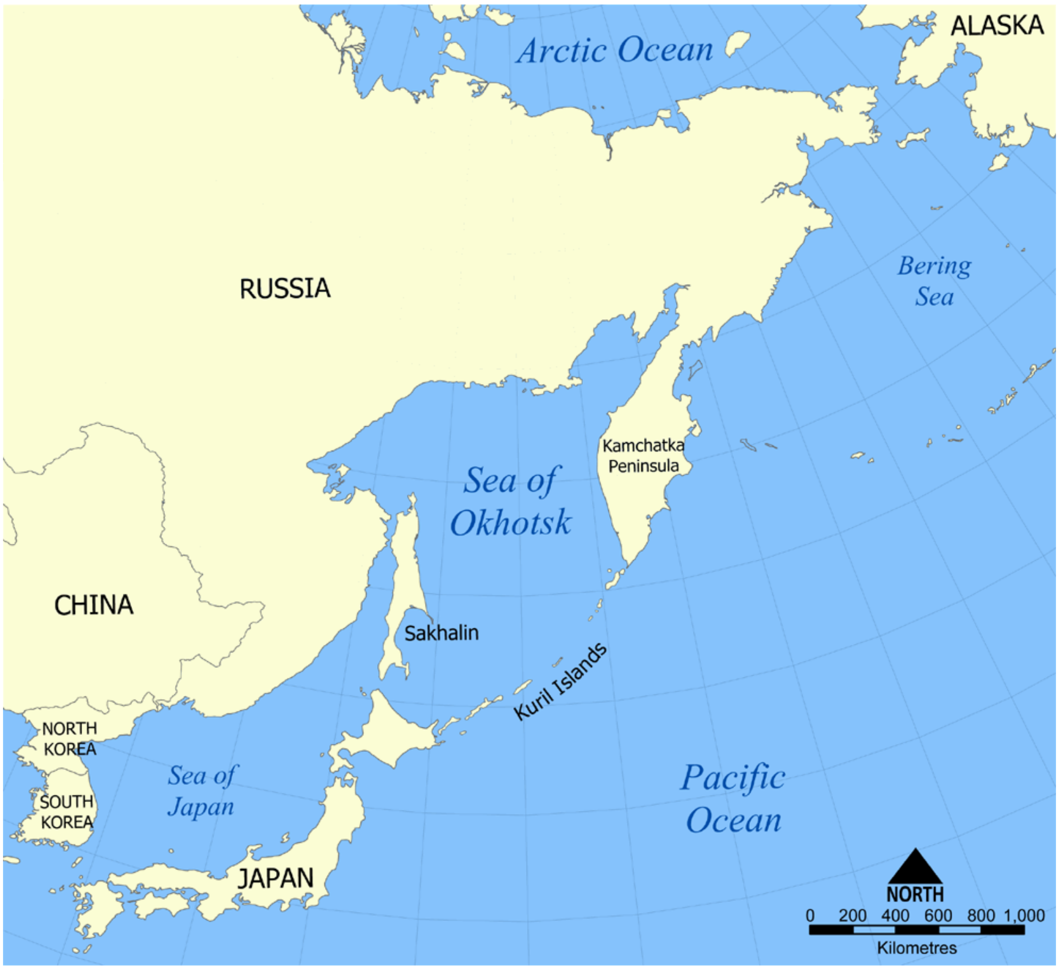


Figure 16.4. Kurile Islands, Kamchatka Peninsula, Aleutian Islands, and Chukotka.Perry-Castañeda Library Map Collection. In Public Domain. http://legacy.lib.utexas.edu/maps

Heizer (1944) presents a report that explores whaling methods across the northern Pacific Rim. The Jomon people of Japan used nets to harvest whales. However, the Ainu of the Kurile Islands and southern Kamchatka Peninsula utilized a dart or lance that had been coated with aconite poison. An individual or small hunting party paddled out to sea and stabbed a whale just one time. The poison eventually killed the whale. Hunters then waited for the dead animal to float ashore. This Ainu method of harvesting whales was later adopted by the Aleutian Islanders and the Alutiiq (Pacific Yupik) on Kodiak Islands. This suggests that the Aleutian Islands facilitated linguistic, cultural and genetic exchange between Alaska and northeastern Asia.

In his 1944 report, Heizer stated that the Koryak of the Kamchatka Peninsula employed a much different method of harvesting whales compared to that employed by the Ainu. The Koryak method utilized a larger hunting party and large boats. The hunters rowed out to sea and stabbed a whale repeatedly with harpoons. The harpoon had a detachable point that affixed a line and float to the whale. Eventually the whale succumbed to the stab wounds and exhaustion. Then it was towed ashore. The Koryaks method later spread to the Chukchi and Asian Eskimos of Chukotka, and then across the Bering Sea, where it was adopted by Alaskan Eskimos along the Arctic Ocean.

8.4. Genetic Evidence of Bi-Directional Flow.

Turning now to the genetic data, in 2010 researchers (Rasmussen et al.) reported that they had sequenced the genome of the so-called "Saqqaq man," a Paleo-Eskimo who died about four thousand ago in northwestern Greenland. Later it was determined that Saqqaq has the Q1a-B143 mutation (Grugni et al. 2019). Q1a-B143 has also been detected in Eskimo remains in northeastern Russia (for additional details see Supplementary Table 16.8: Reference Nos. 18 to 20). Contemporary data also indicates the presence of Q1a-B143 among the Eskimos and Athabaskans, and potentially among contemporary Koryaks and Eskimos of northeastern Asia (see Supplementary Tables 16.15 and 16.16 for additional details).

From a genetics perspective, the Saqqaq data underscores the idea that Amerindians are genetically distant from Eskimo-Aleuts. Eskimo lineages include those that are downstream from Q1a-F1096 as well as Q1b-M346. Amerindian lineages, on the other hand, are downstream from Q1b-M346 and do not include Q1a-F1096 lineages (see Supplementary Figure 16.1: Part A).

At this point we dig deeper into the phylogeny of the Q1a-F1096 mutation in order to understand the evolutionary history of the Q1a-B143 mutation. Downstream from Q1a-F1096 is Q1a-F746, which splits into Q1a-M120 and Q1a-B143. Q1a-M120 eventually became a minor East Asian lineage (see Supplementary Table 16.17), whereas Q1a-B143 became a significant lineage among the Eskimos and Koryaks.

The data support the evolution of Q1a-B143 in northeastern Asia. First, according to Wei et al. (2018), the Q1a-M120 and Q1a-B143 mutations split from Q1a-F746 about 15 thousand years ago. Second, the Kolyma-1 sample, which was taken from a man who 10 thousand years ago in northeastern Siberia, belongs to Q1a-M120 (Sikora et al. 2018; Supplementary Table 16.8: Reference No. 21). Finally, the discovery of Q1a-B143 in the Saqqaq man remains supports the archeological record that places the origins of the Arctic Small Tool expansion in northeast Asia, and the termination of the expansion in Greenland (see Section 7).

At this point we transition to the Q1a-B277 mutation. As reflected by Supplementary Figure 16.1: Part A, Q1a-M25 is downstream from Q1a-F1096. Q1a-L712 is downstream from Q1a-M25. Q1a-B277 is downstream from Q1a-L712. Asian origins of Q1a-B277 are supported by the contemporary distribution of the Q1a-M25 mutation (see Supplementary Table 16.17) which includes Turkmen in Afghanistan and Mongols in western Mongolia. Data for Q1a-B277 comes from ancient DNA samples (Flegontov et al. 2017; Supplementary Table 16.8: Reference Nos. 13 to 17). The data includes Ust'-Belaya man, an Eskimo who died 4.2 thousand years ago in Chukotka, Russia. Additionally, the Q1a-B277 data comes from Eskimo and Athabaskan remains that were discovered in Alaska.

The Q1a-B143 and Q1a-B277 mutations reflect geneflow from northeastern Asia into Alaska that began at least four thousand years ago. However, geneflow across the Bering Sea was not unidirectional. Rather Q1b-B34 data suggest that geneflow also occurred in the opposite direction, from Alaska to northeastern Asia. In order to better understand the evolutionary history of this mutation, the reader is directed to Supplementary Figure 16.1: Part C. As shown by the figure, Q1b-M3 splits into Q1b-M848 and Q1b-Y4276. Downstream from Q1b-Y4276 is Q1b-B34. Grugni et al. (2019) outline several salient points about the Q1b-Y4276 mutation. First, it is distributed from Siberia to South America. Second, the mutation evolved in the Americas about 9.3 thousand years ago. Third, the Q1b-B34 downstream mutation represents a back migration of Native Americans into northeastern Asia. Fourthly, based on dating estimates obtained from Koryaks, the back migration occurred about five thousand years ago. Finally, ancient DNA supports the back migration (see, also, Supplementary Table 16.8: Reference Nos. 5, 7, and 8).

Section 9: Problematic Whole Genome Perspective of Bi-Directional Flow.

The term "whole genome" reflects attempts to use autosomal data as a tool for deciphering human genetic history. As detailed previously in the first chapter, autosomal research utilizes alleles rather than mutations as a genetic tool, whereas mtDNA and Y-chromosome data utilize mutations that are found on non-recombinant regions of the human genome. As such, analysis of the autosomal data requires complex statistical analysis to overcome the reshuffling of genetic cards that occurs as the result of recombination. mtDNA and Y-chromosome data overcome this problem as they gathered from non-recombining regions of the genome. Y-chromosome data become the tool of choice because the larger size of this locus provides a much more detailed picture of genetic variation.

Two reports (Flegontov et al. 2016a; Flegontov et al. 2016b) take the position, based largely on complex statistical analysis of whole genome data, that Paleo-Eskimos came from the Siberian interior and spoke a proto-Yeniseian language. According to the reports, they crossed the Bering Sea into Alaska. Finally, the reports suggest that contemporary Athabaskans represent the linguistic and genetic relics of this Paleo-Eskimo migration. Contemporary Alaskan Eskimos, on the other hand, are the ancestors of Neo-Eskimos that occupied the Alaskan coastline about a thousand years ago.

Wei et al. (2018) make a compelling argument about the limitations of genomic and mtDNA data as tools for deciphering the genetic history of Native Americans. According to the researchers, Y-chromosome single nucleotide polymorphisms are the tool of choice. Extending their argument further, triangulated Y-chromosome based modeling provides a far more robust transparent methodology for deciphering the language prehistory of Athabaskans and Eskimos. A synthesis of Y-chromosome, archaeological, climatological, and linguistic data, as provided in Section 8 above, fails to place the origins of Paleo-Eskimos in Central Siberia. Moreover, these data fail to support a macro-family linguistic relationship for Yeniseian and Eyak-Athabaskan.

Section 10. Problematic Models of Native American Origins.

10.1. Overview.

At this point a discussion of problematic of Native Americans origins is presented. The goal of this discussion to provide linguists with information they need for future models that present a triangulated Y-chromosome perspective of Native American languages.

10.2. The Genetic Ancestry of Native Americans.

Genomic reports have surfaced that posit ancient Asians as the genetic ancestors of contemporary Native Americans. Raghavan et al. (2014), for example, report data for the so-called Mal'ta boy. He was a two-year-old child who died along the banks of Belaya River in southern Siberia about 24 thousand years ago. The study suggests that Native Americans derive a significant part of their genetic ancestry from this individual. Another example is the report from Sikora et al. (2018). They present data for Kolyma-1, an individual who died ten thousand years ago in northeastern Siberia. The study suggests that Native Americans derive part of their genetic ancestry from this man.

The genetic history of Native Americans is a legitimate research question. However, it should be emphasized that genes do not define ethnicity. Rather, the question of identity is a matter that Native Americans should define for themselves. Thus, as an ethical matter, it is important for researchers to differentiate genetic history from ethnicity. Mal'ta and Kolyma-1 might be genetic ancestors but they are not cultural ancestors.

Apart from ethical considerations, researchers should question whether the data support the position that Mal'ta and Kolyma-1 are, in fact, genetic ancestors of Native Americans. This suggests an over-expansive interpretation of the data. A more conservative treatment of the Mal'ta data suggests that he merely represents part of the genetic inventory (or genome) of those living in south-central Siberia at the time of the Last Glacial Maximum. Kolyma-1, on the other hand, merely represents part of the genome of northeastern Siberia at the beginning of the Holocene.

10.3. Beringian Standstill.

Those that explore the archeology and genetic history of Native Americans will certainly encounter the term "Beringian standstill." This model was initially proposed by Tamm et al. in their 2007 report. The researchers questioned the speed of the initial human migration wave into the Americas. They considered whether it was a rapid "direct colonization" event or, alternatively, if humans congregated in a refugium near the Bering land bridge before migrating into the Americas. Based on a comparison of Asian and New World mitochondrial DNA lineages, the researchers favored an "incubation" period, meaning that the first humans in America underwent genetic isolation for up fifteen thousand in a northeastern Asian refugium before migrating over the land bridge into Alaska.

The Beringian standstill hypothesis cannot be defended with archeological data (e.g. Buvit and Terry 2016; Potter et al. 2018). One huge problem is fuel for a fire (Hoffecker et al. 2014). The hypothesis is also problematic from a Y-chromosome perspective. As suggested by Wei et al. 2018, Y-chromosome diversity downstream from Q1b-L54 posits a rapid human migration from south-central Siberia beginning about 16 thousand years ago. Finally, as detailed in Section 4, Beringian standstill is inconsistent with climatological record.

10.4. Pre-Clovis Human Migrations into the Americas.

Section 3 of this present chapter provides a robust settlement model of the Americas that is well supported by a synthesis of archeological, climatological, genetic and Y-chromosome data. In summary, *Homo sapiens* crossed over the Bering Sea into Alaska roughly fifteen thousand years ago. By around 14 thousand years ago, they had migrated southwards through an Ice-Free Corridor onto the Great Plains of North America. Around ten thousand years ago, they arrived in South America.

The Clovis culture can be securely dated to around 13 thousands years ago (see Section 4). Archeological studies periodically surface that present evidence of "pre-Clovis" migrations. Bourgeon, Burke and Higham (2017) report, for example, human presence in North America by around 24 thousand years ago. This is based on cut marks on bones found at the Bluefish Caves site in the Yukon of Canada. According to researcher, the cut marks were clearly made with human-made tools. Yet the study does not account for the possibility that scavengers made the cut marks ten thousand years after the animals had died. Conditions are such in the Arctic that animal remains are well preserved in ice for thousands of years. Dillehay et al. (2015) presents another example of pre-Clovis migrations. Based on artifacts found at the Monte Verde archeological site, they suggest that humans arrived in southern Chile around fourteen thousand years ago. However, the use of artifacts to date human presence in an area can be problematic. One has to use organic matter around the site to provide carbon-14 results, and sometimes this provides a poor correlation with human occupation. Dating results from human remains provide the most robust estimates for human occupation. Thus, Monte Verde dating estimates might be problematic because human remains have not been found at this location.

"Pre-Clovis" represents a controversial archeological debate that fails to advance an understanding of Native American languages. We suggest that Pre-Clovis is also a moot point. Pre-Clovis arguments correlate Clovis with the first Native Americans. Very recent archeological and genetic evidence changes this long-standing assumption. Clovis simply represents the terminal end of a long Pleistocene hunter cultural tradition that began 45 thousand years ago in northern Eurasia. The Native American cultural tradition, on the other hand, began about 11 thousand years ago when the mammoths became extinct.

10.5. The Polynesians.

Campbell (1997: 261-262) provides numerous "far-fetched" macro-family proposals that have surfaced in Native American historical linguistics: Amerindian and Basque; Na

Dené and Mongolian; Mayan and Turkic; Quechua and Tungusic. His list of "far-fetched" proposals also includes Native American languages and Austronesian.

Unfortunately, despite a total absence of genetic, linguistic, and archaeological evidence, the rumors of a prehistoric Polynesian migration into the Americas continue to circulate. The Polynesian rumor was once driven by a branch of archaeology called cephalometry, the study and measurement of the head. For example, a study from 1996 (Neves et al.) proposed that Polynesians were among the founding populations of the Americas based on craniometric measurement of fifty-three skulls. Another more famous example is "Kennewick man." He died about eight thousand years ago near Kennewick, Washington in the United States. For several years his remains were associated with Polynesian or Ainu ancestry based on skull measurements (e.g. Taylor, Smith and Southon 2001). However, as reported by Rasmussen et al. in 2015, Kennewick man belongs to Q1b-M3, a Native American lineage.

Genomic studies (Skoglund et al. 2015; Moreno-Mayar et al. 2018) have re-ignited the Native American-Polynesian rumor by reporting an Australasian component among Native American. The statistical modeling developed by Moreno-Mayar (2018) utilized an Andaman Islander with the P-P295 upstream mutation as their Polynesian "proxy." It should be noted that P1-M45 mutations are found in South Asia (Gazi et al. 2013) and Island Southeast Asia (Karafet et al. 2015). However, the same mutation was part of the genetic inventory of Northern Eurasia during Marine Isotope Stage 3. The P1-M45 was recently found in Paleolithic remains from Siberia, the Yana-1 man who died 32 thousand years ago (Sikora et al. 2018). Rather than a Polynesian signal, Skoglund et al. (2015) and Moreno-Mayar et al. (2018) probably detected a North Eurasian genetic signal among Native Americans, which is expected.

10.6. Solutreans.

Among the more dubious models surrounding the origins of Native Americas (Campbell (1997: 90-93) is that they came from the lost island of Atlantis. Closely related to this rumor is the Solutrean hypothesis, that Clovis is a continuation of the European Solutrean cultural tradition that ended roughly 17 thousand years ago. Needless to say this hypothesis is not consistent with mainstream archaeological opinion (see Straus, Meltzer and Goebel 2005). Part of the problem is geographic distance. The lack of unequivocal Solutrean artifacts in the Americas is also problematic.

Despite the absence of archeological data, the Solutrean hypothesis has nevertheless resurfaced because of recent genetic data, the Q1b-L804 mutation. At this point the reader is directed to Supplementary Figure 16.1: Part C. As shown by the figure, Q1b-M3 and Q1b-L804 are sister clade mutations downstream from Q1b-M930. As previously discussed, Q1b-M3 is a Native American signature lineage. Q1b-L804, on the other hand, is found in northeastern Europe where it attains a very small frequency among the men of this region. Given the close phylogenetic relationship between Q1b-M3 and Q1b-L804, Wei et al. (2018) felt compelled to warn researchers that Q1b-L804 does not support the Solutrean hypothesis. Rather, the genetic data support a rapid diversification of Q1b-L54 around 16 thousand years ago in south central Siberia. Q1b-L804 and Q1b-L330 remained in the region. Q1b-M3 and Q1b-Z781 are the relics of Q1b-L54 diversification in the Americas. Such a scenario is consistent with archaeological evidence (see Section 4).

Section 11. Conclusions.

The first humans that migrated into the Americas belonged to a cultural tradition that began 45 thousand years ago in Northern Eurasia: the Pleistocene mammoth hunters. Eleven thousand years ago, when the Holocene began, warmer weather caused a worldwide extinction of the mammoths. At this point, the Pleistocene mammoth-hunter tradition ended in Northern Eurasia and North America. Those that adapted to Holocene climate change in the New World became the Native American cultural tradition. This cultural tradition marks the beginning of linguistic diversity in the Americas.

This incredible diversity of indigenous languages in the Americas correlates well with the diversified subsistence strategy adopted by the Native Americans. They exploited regional resources. These resources brought opportunities and imposed constraints. In some areas, foraging and language evolved *in situ*. The Pacific Coast Indians are an example. However, some foraging cultures, such as the Athabaskans, migrated. In some areas, agriculture and language co-evolved. The co-evolution of agriculture and Quechuan in the Central Andes is an example. However, one finds evidence of language-farming expansions. Arawakan provides a solid example. Finally, some cultures, such the Numic peoples, abandoned agriculture and returned to foraging.

Amerinds languages evolved directly from the Native American cultural tradition that formed at the beginning of the Holocene. Their evolution remained undisturbed from outside influence until 1492. Eskimos-Aleut and Eyak-Athabaskan, on the other hand, were potentially shaped by contact with the indigenous people of northeastern Asia over the last four thousand years. Unfortunately, the paucity of genetic data prevents us from knowing more about the prehistory of these language families. For example, the geographic origins of Eskimo-Aleut remain elusive. How and when these languages expanded also remains a mystery. Similarly, it is difficult to defend or reject Greenberg's proposal that Eyak-Athabaskan stems from a second migration into the America. We simply need more high-resolution Y-chromosome data from contemporary populations.

Chapter 17: Haplogroup R-M207.

Section 1. The Contemporary Distribution of Haplogroup R-M207.

Haplogroup R-M207 has two main divisions within its phylogeny, R1-M173 and R2-M479 (see, also, Supplemental Figure 17.1). The R2-M479 branch is found mostly among populations living in South Asia. The R1a-M420 and R1b-343 mutations define the two main divisions within R1-M173. R1a-M420 mutations are found in Scandinavia, Eastern Europe, the Baltic Region, South Asia, Central Asia and Northern Eurasia. R1b-M343 mutations are mostly found in Western Europe and the Sahel region of Africa.

Section 2. The Evolution of Haplogroup R-M207.

Around 29 thousand years ago Marine Isotope Stage 3 ended. This marks the beginning of Marine Isotope Stage 2. Shortly thereafter, roughly 27 thousand years ago, the Ice Age glaciers reached their maximum southern extension across Eurasia. The literature refers to this event as the Last Glacial Maximum (see Clark et al. 2009 for a more detailed discussion). Within this climatological context, haplogroups R-M207 and Q-M242 diverged from the P1-M45 mutation roughly 30 thousand years ago (Poznik et al. 2016: Supplementary Table 10). See, also, Supplementary Figure1.1 from the first chapter.

Haplogroups R-M207 and Q-M242 represent the genetic signature of those who survived the Last Glacial Maximum. Around the time of the Last Glacial Maximum, glaciation pushed human populations southwards across Eurasia into what the literature describes as "refugia" (e.g. Gavashelishvili and Tarkhnishvili 2016). The ancient DNA data, along with the archeological and climatological data, suggest that haplogroups R-M207 and Q-M242 co-evolved in a south-central Siberian refugium in the Altai-Sayan region, where the borders of contemporary Russia, Kazakhstan, China and Mongolia converge. Ancient DNA evidence comes from the so-called "Mal'ta boy." He died about 24 thousand year ago near Lake Baikal, which is the general vicinity of the Altai-Sayan region. As previously noted in Chapter 16: Section 3, researchers place the child's Y-chromosome somewhere near the root of the R-M207 haplogroup, where haplogroups R-M207 and Q-M242 diverge from P1-M45 (see, also, Raghavan et al. 2014).

The climatological record suggests that the Altai-Sayan region and Eastern Europe were spared from the extreme glaciation that had occurred in Western Europe (e.g. Velichko et al. 2009; Binney et al. 2016). Cold and arid climatic conditions within a "hyperzonal open-type" landscape supported large mammal food resources for human populations across a vast geographical expanse. This so-called "hyperzone" included South-Central Siberia and much of Eastern Europe about three hundred kilometers south of the ice glaciers, which equates to around 50 degrees north for both regions (Velichko et al. 2009). According to Kuzmin (2008), because of the large mammal food resources that were supported by the hyperzone, human population density in South-Central Siberia remained relatively stable. Furthermore, the researcher asserts that other factors contributed to successful cold weather adaptation for those living in this region during the Last Glacial Maximum: micro-blade tools, suitable dwellings and clothing, and the availability of bones for fuel.

Meanwhile the glacial ice sheet in Western Europe pushed much further south during the Last Glacial Maximum, almost to 40 degrees north. Kuzmin (2008) attributes greater glaciation in Western Europe to a shift in the Atlantic storm track that brought more moisture to the region. As a result, Scandinavia and much of Western Europe became depopulated and human populations retreated to the Iberian Peninsula. Interestingly, in contrast to refugia in South-Central Siberia and Eastern Europe, human population density within the Iberian refugium may well have declined as the result of a decline in reindeer populations (Jochim et al. 1999; Morein 2008).

The comparison of Iberian and Siberian population densities during the Last Glacial Maximum, as presented in the previous paragraph, potentially explains why NO-M214 and C-M130 mutations disappeared from the European gene pool. As previously discussed in Chapter 6: Section 4 and Chapter 14: Section 2, human remains and ancient DNA data suggest that both mutation were very much part of the European gene pool during Marine Isotope Stage 3. Thus, a population bottleneck may explain their demise, and this explains why haplogroup I-M170 remains the sole uncontested founder lineage among contemporary Europeans. Taking this a step further, greater reproductive success during the Last Glacial Maximum partially explains the ubiquitous presence of R-M173 mutations among the contemporary population of Eurasia.

Section 3. The Expansion and Diversification of R1-M173.

The R1-M173 mutation arose about 28 thousand years (Poznik et al. 2016: Supplementary Table 10). The contemporary distribution of R2a-M124 mutations (see Section 10), along with the climatological and archeological evidence, as presented previously in Section 2, suggests that this occurred in south-central Siberia. Then about 23 thousand years ago, R1a-M420 and R1b-M343 diverged from R1-M173 (Poznik et al. 2016: Supplementary Table 10). Underhill et al. (2015) suggest, based on analysis of contemporary Y-chromosome data, the split occurred in the Middle East in the vicinity of Iran. However, archaeological and ancient DNA evidence suggests the East European Plain (see Figure 17.1 below) in the vicinity of Kiev in the Ukraine. Archaeological support for the European Plain stems from Abramova et al. and their 2001 report of Upper Paleolithic sites in the middle Dnieper River basin. These sites date from about 12 to 25 years thousand ago, and as such, they extend back in time to the Last Glacial Maximum. Furthermore, these sites are found along the same latitude as the Altai-Sayan refugium. Finally, ancient DNA Y-chromosome data from the Baltic region and Eastern Europe support this idea (see Supplementary Tables 17.1 and 17.2).

Section 4. Diversification of R1b-M343 on the East European Plain.

Dolukhanov (2009) provides a discussion of hunter-gatherer cultures at the time of the Paleolithic-Mesolithic transition of the Eastern European, about 12 thousand years ago. From a climatological perspective, this transition marks the beginning of the Holocene. Connected to this change in climate was the southward expansion of human populations from northeastern Europe to Southeastern Europe. Taking this a step further, ancient DNA, phylogenetic relationships and the contemporary distribution of modern-day Y-chromosome variation support the following idea: initial diversification of R1b-M343 variation occurred in northeastern Europe and southeastern Europe on the East European Plain. Ancient DNA support for this position comes from Supplemental Table 17.1. As shown by the table, at the onset of the Holocene R1b-M343 mutations had already spread across the East Europe Plain along a north to south axis.

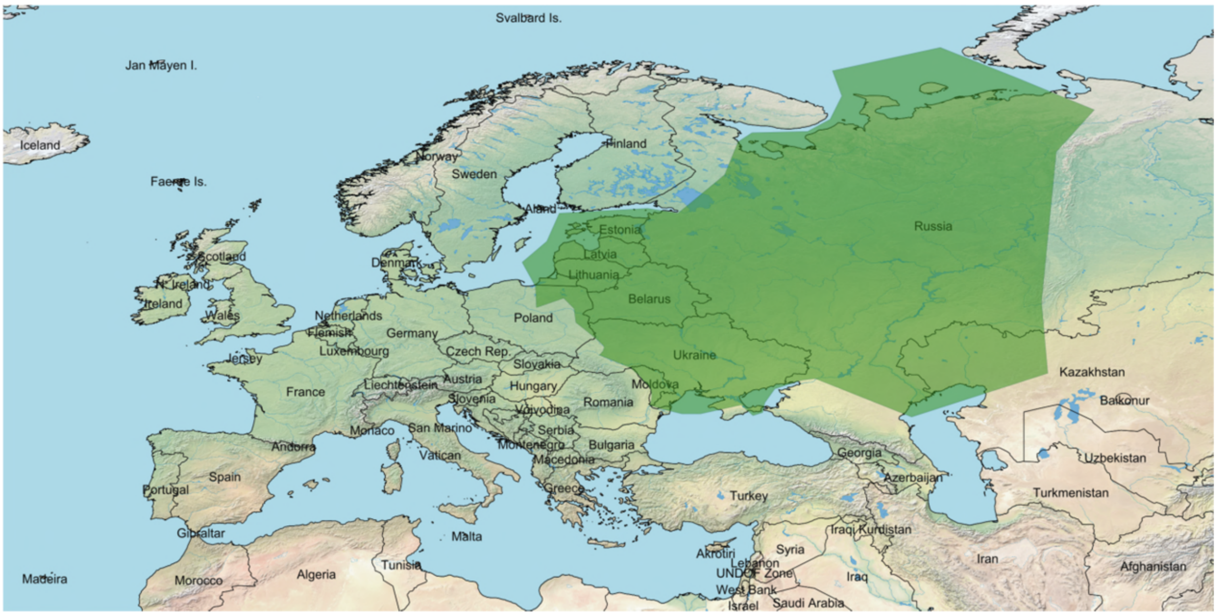


Figure 17.1. The East European Plain.

At this point the reader is invited to review Supplementary Figure 17.3. The R1b-M335 marker was found in eight thousand-year-old remains in Latvia (Jones et al. 2017). Given the unique position of R1b-M335 within R1b-M343 phylogeny, we have ancient DNA evidence that support diversification of R1b-M343 in Northeastern Europe. Southeastern European diversification of R1b-M343, on the other hand, is supported by the contemporary distribution of R1b-M73 (see Supplementary Table 17.3), R1b-V88 (see Section 7) and R1b-CTS1078 (see Section 8).

Section 5. The Expansion of R1b-M269 into Western Europe.

R1b-M269, like R1b-M335, evolved in northeastern Europe. Around the onset of the Holocene, R1b-M269 expanded westward across the North European Plain (see Figure 17.2 below) into Western Europe. The mutation was carried by reindeer hunters. Secondary expansions then carried the mutation into Scandinavia, the British Isles, Iberia and the Mediterranean.

Turning now to the archaeological record, during the Last Glacial Maximum human populations in Western Europe retreated to southern refugia including the Iberian Peninsula (for additional details, see Section 3). Based on radio-carbon estimates, about 14 thousand years ago people and reindeer began their expansion northwards (Housley et al. 1997). The glaciers slowly retreated. Tundra then appeared in previous glaciated area. The tundra eventually succumbed to the forests, and the tundra line gradually retreated to its present location above the Arctic Circle. The reindeer followed the retreating tundra line, and hunter-gatherers followed the reindeer (see Sommer et al. 2014 for additional details). By around 12 thousand years ago, the reindeer arrived in Denmark, and 9 thousand years ago, they disappeared into the arctic region of Scandinavia (Aaris-Sorensen et al. 2007).

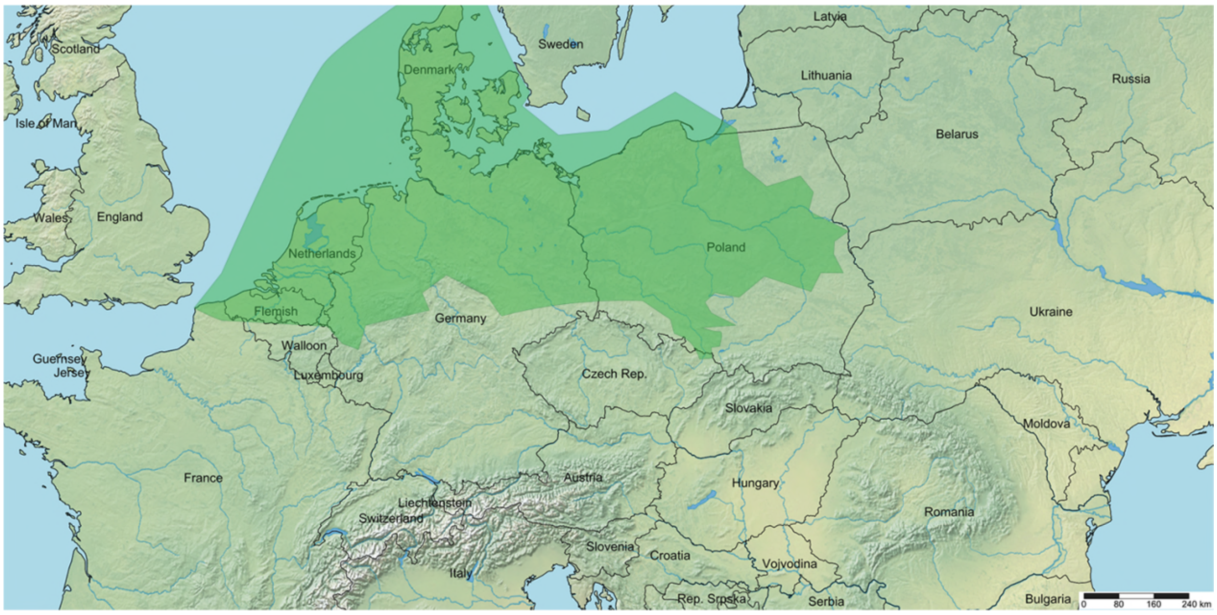


Figure 17.2. The North European Plain.

The archaeological evidence reflects that around the beginning of the Holocene (12 thousand years ago) almost all of the mega-faunal food resources of northeastern Europe, such as mammoths, had disappeared (e.g. Puzachenko and Markova 2019). The last remaining mega-faunal resource at this time was reindeer (see Dolukhanov 2009: 32 and the discussion of the Swiderian culture). Hunter-gatherers on the East European Plain were then drawn to Western Europe because of the abundance reindeer in this region and the demise of mega-faunal resources elsewhere.

Based on genetic evidence, at the onset of the Holocene hunter-gatherers intercepted the northward migration of reindeer in Western Europe and Scandinavia along two different trajectories. As explained previously in Chapter 9, haplogroup I-M170 represents the genetic signature of the south-north trajectory. The R1b-L51 mutation, on the other hand, represents the genetic signature of the east-west trajectory. The R1b-L51 mutation is a downstream variant of the R1b-M269 mutation. According to Myres et al. (2011), this mutation evolved in Western Europe.

Sometime during the early Mesolithic, R1b-U106 and R1b-S116 diverged from R1b-L51. R1b-U106 reflects genetic diversification of R1b-L51 on the North European Plain, the British Isles and Scandinavia (see Myres et al. 2011 and Supplementary Table 17.4). The R1b-S116 mutation reflects diversification of R1b-L51 in Iberia (Valverde et al. 2016).

R1b-S116 has three informative downstream mutations: R1b-DF27, R1b-U152, and R1b-M529. Valverde et al. (2016) suggest based on their analysis of the genetic evidence that these three mutations diverged from R1b-S116 on the Iberian Peninsula around 12 thousand years ago. They also take the position that the R1b-DF27 mutation remained on the peninsula. R1b-U152 on the other hand expanded eastwards onto the Italian Peninsula and then northwards, perhaps through the Alps into Germany. Like R1b-U152, R1b-M529 also expanded out of the Iberian Peninsula. However, this mutation expanded along a different trajectory to the British Isles. This expansion model is supported by the contemporary distribution of all three mutations (see Supplementary Tables 17.5 to 17.7) and the observation that the highest frequency of S116 mutations is found in Spain and Portugal (Valverde et al. 2016).

Dating estimates provided by Solé-Morata et al. (2017) suggest an expansion of R1b-U152 and R1b-M529 from the Iberian Peninsula around four thousand years ago. In contrast to this estimate, coalescent times provided Myres et al. (2011: Table S2) suggest the beginning of the Neolithic transition on the Iberian Peninsula. These estimates from Myres et al. (2011) are more consistent with the archeological record. According to Martins et al. (2015) agriculture on the Iberian Peninsula dates to around 7.5 thousand years ago. This is based mostly on radio-carbon results taken from the remains of domesticated sheep and goats, an exceptionally reliable data source for delineating the Neolithic/Mesolithic transition. Their data further suggest that the Neolithic transition drove rapid population growth within the region. Thus it seems that population pressure during the Early Iberian Neolithic resulted in an expansion of R1b-U152 and R1b-M529 from the peninsula.

One interesting observation from recent Iberian studies (Valverde et al. 2016 and Solé-Morata et al. 2017) is that the highest frequency of R1b-DF27 mutations is found among the Basque people. This a salient point for linguists because many researchers consider the Basque language isolate a linguistic relic of Europe from a time that predates the arrival of Indo-European languages. In Chapter 9: Section 7, genetic support was provided for this position based on data from the I2a-M26, and C-M130 mutations. The R1b-S116 mutation, and its arrival in the Basque region around the beginning of the Holocene, provides additional support.

Section 6. R1b-S116 and Celtic Languages.

In contemporary Europe languages belonging to the Celtic branch of Indo-European are spoken in Ireland, the United Kingdom and the Brittany region of France. *Ethnologue* (2018) classifies the extant Celtic languages into a single “Insular” branch. The Insular branch is subdivided into Brythonic and Goidelic. Brythonic consists of Breton and Welsh. Irish Gaelic and Scottish Gaelic are the Goidelic languages. In prehistoric Europe, Celtic languages had a much broader distribution extending from the Atlantic Ocean to Asia Minor. The classification of the pre-Roman Celtic languages not only included the Insular branch but also Continental Celtic branch of languages that are now extinct. Examples include Celtiberian on the Iberian Peninsula, Gaulish in France, and Leponic in Northern Italy and Switzerland.

As noted previously in Section 5 (above), the R1b-DF27, R1b-M529, and R1b-U152 are downstream from the R1b-S116 mutation (see, also, Supplementary Figure 17.3). As shown by Supplementary Tables 17.5 to 17.7, the available data for downstream variants of R1b-S116 are rather limited. Nevertheless, the available data strongly link R1b-U152 with Continental Celtic, and R1b-M529 with Insular Celtic.

Section 5 (above) presents a synthesis of several different data sources that include high resolution Y-chromosome mutations. This synthesis of data suggests that an expansion of farming from southwestern Asia brought proto-Celtic to the Iberian Peninsula. Celtiberian represents the diversification of proto-Celtic this region. The other Continental Celtic languages, such as Gaulish and Lepontic, as well Insular Celtic, represent a co-expansion proto-Celtic-speaking farmers from Iberia and the later diversification of proto-Celtic in other regions of Europe. This model of proto-Celtic origins, which utilizes Y-chromosome mutations, is strikingly similar to one that utilizes mitochondrial DNA (see McEvoy et al. 2004).

The triangulated Y-chromosome based model of proto-Celtic conflicts with a recent paleo-genomic study. Cassidy et al. in their 2016 study conducted paleo-genomic analysis of four ancient remains from Ireland. The remains date to the Neolithic/Bronze transition on the island, roughly four thousand years ago. According to the researchers, the R1b-M529 mutation and Celtic languages arrived in Ireland during the Bronze Age. They also place the origins of R1b-M529 and Celtic languages on the Central Eurasian steppes and the expansion of the Yamnaya culture. This assertion is problematic as the genetic signature of these steppe nomads belong to a mutation that is phylogenetic distant, the R1b-CTS1078 mutation (see Supplementary Figure 17.3 and the discussion below in Section 8). Thus, the conclusions rendered by Cassidy et al. (2016) illustrate the potential pitfalls of using statistical methods to model the prehistory of language as well as the potential benefits of utilizing triangulated Y-chromosome based modeling.

Section 7. The Expansion of R1b-V88 into Africa.

The highest frequencies of the R1b-V88 mutation are found in the Sahel region of Africa, a zone that divides North Africa and Sub-Saharan Africa (see Supplementary Table 17.8). The archaeological and genetic data suggest that the mutation evolved in southeastern Europe during the Mesolithic. Support for this position comes from a recent study (D'Atanasio et al. 2018) that reports the evolution of R1b-V88 about 12 thousand years ago in Europe. According to the study, about eight thousand years ago the mutation was carried southwards into northeastern Africa. From this location, the mutation expanded southwards into the Sahel.

The R1b-V88 expansion into Africa stands as genetic relic of the North African Mesolithic and the so-called "humid phase." The previous discussion of the A-M13 mutation (Chapter 2: Section 3) reports that about 10 thousand years ago the climate of northern Africa underwent a dramatic transformation. Holocene climate change brought monsoon rain to the region. The Sahara Desert became a savannah with numerous lakes and rivers. Within the complex system of rivers and lakes, hippos, crocodiles and fish proliferated. Archeological evidence demonstrates that human hunted these animals.

Abundant food resources appear to have pulled the R1b-V88 mutation into Northern Africa. Several lines of evidence support such a position. Interestingly, the oldest R1b-M343 sample comes from the Villabruna remains found in the Dolomite Mountains region of Northern Italy (see Supplemental Table 17.1 and data for the Mediterranean region). The remains are from a man who died around 14 thousand years ago. Researchers determined that he belongs to R1b1-L278, and as such, he becomes a potential genetic ancestor of R1b-V88 populations. Partial support for this position comes from the previous discussion in Section 5, which suggests that Mesolithic R1b-V88 mutations in Italy were replaced by the Neolithic R1b-U152 mutation. A similar position was taken by D'Atanasio et al. (2018). The same study also reports that R1b-V88 mutations among contemporary Sardinians are older than African Rb-V88 mutations. Furthermore, the Sardinian samples date to the beginning of the North African humid phase.

Kuper and Kröpelin in their 2006 study suggest that during the humid phase the Nile River was a marshland and people were not able to live in this region. As such, a Mesolithic expansion of R1b-V88 via the Middle East and northeastern Africa would have been difficult. On the other hand, palaeohydrological data taken from satellite imagery (Drake et al. 2011: Figures 1 and 2) show numerous river systems that were present in North Africa during the humid phase. Taking this a step further, hunter-gatherers potentially made a water crossing from Sardinia to the North Africa. From Mediterranean coastline of North Africa, these rivers could have facilitated a southward migration to the Mesolithic food resources in the vicinity of Lake Chad.

Chadic is a branch of the Afro-Asiatic language family. In a 2010 study, Cruciani et al. suggested that R1b-V88 is a genetic signature for Chadic languages based on the high frequency of the mutations among the Chadic-speaking populations of Africa. Such a position seems problematic for several reasons. Haber et al. (2016) disputed the conclusion from Cruciani et al. (2010) asserting that the oldest R1b-V88 mutations are found among the Laal-speakers, a language isolate. Data for contemporary African populations (see Supplemental Table 17.9) also show that R1b-V88 is found among speakers of Semitic and Berber languages, which are also branches of the Afro-Asiatic family. Additionally, R1b-V88 is well represented among several populations that speak either a non-Bantoid Niger-Congo languages or Nilo-Saharan languages. Finally, the E-M34, E-M81 and J1-M67 mutations, along with the archaeological data, suggest that Afro-Asiatic expanded into Africa from the Middle East during the Neolithic (see Chapter 5: Section 3 and Chapter 10: Section 3), whereas R1b-V88 is a Mesolithic relic among African populations.

Section 8. The DNA of Steppe Nomads.

As the result of palaeogenomic studies, Bronze Age steppe nomads of Eastern Europe have taken center stage in the debate surrounding the prehistory of Indo-European languages (see Section 11 of this present chapter). According to the ancient DNA data, the genetic signature of these nomads is the R1b-CTS1078 mutation (see Supplemental Table 17.10). Based on contemporary data (see Supplementary Table 11), and phylogenetic relationships (see Supplementary Figure 17.3), the R1b-CTS1078 mutation appears to have evolved in southeastern Europe. Frequency results from Supplementary Table 17.11 clearly rule out a massive migration of the steppe nomads into Western Europe during the Bronze Age.

Section 9. Diversification of R1a-M420 on the East European Plain.

As previously explained in Section 4, R1a-M420 and R1b-M343 diverged from R1-M173 about 23 thousand years ago somewhere on the East European Plain. Like R1b-M343 (see Section 3), the initial diversification of R1a-M420 variation occurred in northeastern Europe and southeastern Europe. R1a-M420 has two main downstream variants, R1a-Z282 and R1a-Z93 (see Supplementary Figure 17.2). Based on contemporary population data (see Supplementary Table 17.12) it appears as though the R1a-Z282 mutation represents diversification of R1a-M420 in northeastern Europe. Limited support for this position comes from ancient DNA data (see Supplementary Table 17.2).

R1a-Z282 has three informative downstream markers: R1a-Z284, R1a-M458, and R1a-M558. R1a-Z284 is confined almost exclusively to Scandinavia where it attains a frequency of around twenty percent among Norwegians (Underhill et al. 2015). Based on coalescent time estimates (Underhill et al. 2015: Table S5) it appears as though the evolutionary history of R1a-Z284 appear to be similar to that of R1b-U106 (see Section 5).

Specifically, both are genetic relics of an early Mesolithic expansion of reindeer hunters.

The R1a-M448 and R1a-M558 mutations extend across the northern East European Plain. Based on contemporary data, R1a-M448 appears to have a higher frequency among West Slavic speakers (see Supplementary Table 17.14) and R1a-M558 seems to peak among East Slavic populations (see Supplementary Table 17.15). Thus, it should be emphasized that these data results should not be associated with the historical Slavic expansion. Rather, language contact during historical times explain the contemporary distribution of Slavic languages (e.g. Brackney 2007).

Contemporary data suggest that R1a-Z93 represents diversification of R1a-M420 mutations in southeastern Europe or perhaps Central Asia (Supplementary Table 17.13). R1a-Z93 is an especially significant component of South Asian populations. Underhill et al. (2015: Table S4) suggest a frequency of about eighteen percent for the region. At this point the reader is invited to examine Supplementary Table 17.16. The table sorts frequency data for R1a-Z93 according to language family or language branch. These data reflect that R1a-Z93 is a significant marker among populations that speak Iranian, Indo-Aryan, Dravidian and Turkic languages. According to Pamjav et al. (2012), R1a-Z93 diverged from R1a-M420 about 10 thousand years ago. These data along with the archeological record support a model that associates the evolutionary history of R1a-Z93 with Holocene climate change. Misra (2001) reports that population density was low during the Upper Paleolithic in India. Arid and cold weather had limited the availability of food resources. The Mesolithic, however, brought monsoon rains. Increased moisture produced more food resources that ultimately drove higher population density.

Section 10. The Evolutionary History of R2-M479.

As noted previously in Section 1, R1-M173 and R2-M479 form the two main downstream divisions of the R-M207 haplogroup. Data from Poznik et al. (2016: Supplementary Table 10) suggest that the R2-M479 mutation evolved roughly 28 thousand years ago, around the time of the Last Glacial Maximum. As suggested in Section 2, this appears to have occurred in South-Central Siberia. Contemporary data for the mutation consists almost entirely of frequency results for the R2a-M124 downstream mutation. Supplementary Table 17.17 reports R2a-M125 frequencies from a regional perspective. As shown by the table, almost all the reported data for the mutation comes from South Asia.

Ancient DNA helps to resolve the evolutionary history of R2a-M124. As shown by Supplementary Table 17.18, R2a-M124 was extracted from a 10 thousand year old sample found at Ganj Dareh in northwestern Iran. This data supports the idea that climate change drove R2a-M124 into South Asia during the Mesolithic, a model that is similar to that of R1a-Z93 (see Section 9 above).

Sengupta et al (2006) suggest that about nine percent of Indian and seven percent of Pakistani males have the R2a-M124 mutation. At this point the reader is directed to Supplementary Table 17.19 which reports R2a-M124 data from a language perspective. The mutation appears to be an especially informative marker for deciphering the population history among those that speak Dravidian and Indo-Aryan languages. Like R1a-Z93, R2a-M124 was present in South Asia before the arrival of Indo-European speaking population.

Section 11. Problematic Palaeogenomic Modeling of Indo-European Languages.

"Palaeogenomic modeling" is sub-specialty found amongst the whole genome studies. This represents an attempt to model human population history by employing statistical analysis of ancient DNA, and more specifically, autosomal DNA markers that are inherited from both parents. Statistical methods are utilized in order to overcome the problem of recombination. This "reshuffling" of genetic traits can distort and erase evolutionary relations that are needed to decipher human population history. Of course, Y-chromosome data avoid this problem as they are gathered from a non-recombining region of the human genome (see Chapter 1: Section 3).

Especially problematic are recent palaeogenomic models of Indo-European languages. In order to explain why this is problematic, it is necessary to provide some background information. The origin and expansion of this language family has historically followed two different models. The so-called *steppe nomad hypothesis* associates the spread of Indo-European with a Bronze Age expansion of steppe nomads from Eastern Europe or Central Asia (e.g. Gimbutas 1997; Anthony 2007). It should be emphasized that this approach is archeologically weak (e.g. Renfrew 1987; Frachetti 2012). Rather, the hypothesis is largely based on phonological reconstructions (for a recent discussion, see Anthony and Ringe 2015).

The alternate model of Indo-European origins, the *early farming dispersal hypothesis*, suggests that this language family evolved in the Middle East and expanded out of the region roughly nine thousand years ago during the Neolithic. This hypothesis was proposed by the archaeologist Peter Bellwood in his 2005 book *First Farmers: the Origins of Agricultural Societies*. Bellwood supports his model with archeological evidence. Additional support comes from the observation that Neolithic expansions explain the prehistory of other large language families, such as Afro-Asiatic, Niger-Congo, Sino-Tibetan, and Austronesian. Taking this a step further, Indo-European is simply not an exception to the rule.

In 2015 Haak et al. published a palaeogenomic study in the journal *Nature* that takes sides in the dispute surrounding the origins and expansion of Indo-European languages. The study endorses the *steppe nomad hypothesis* of Indo-European language origins. Based on their statistical analysis of ancient autosomal markers, they assert a "massive" expansion of steppe nomads associated with the prehistoric Yamnaya culture of Eastern Europe and Central Asia. The study times the expansion to the Bronze Age, roughly four thousand years ago. According to the study this expansion was so massive that the steppe nomads replaced seventy-five percent of the pre-existing farmer genes in Central Europe. The study further asserts that the R1a-M420 and R1b-M343 mutations are the Y-chromosome relics of this massive invasion of steppe nomads. The basis of this assertion is not exactly clear. However, it is inconsistent with studies that map the evolutionary history of both markers with phylogenetic analysis (Myres et al. 2011; Underhill et al. 2015).

The position taken by Haak et al. (2015) has been endorsed by several subsequent palaeogenomic studies (e.g. Allentoft et al. 2015; Jones et al. 2015; Cassidy et al. 2016; Jones et al. 2017; Olalde et al. 2018; Narasimhan et al. 2019). Based on the fact that these studies are often published in respected science journals, and the fact that they often cited in the news media, linguists might also be tempted to endorse palaeogenomic modeling to decipher language prehistory.

Sections 1 to 11 (above) provide a triangulated Y-chromosome perspective for the prehistory for Indo-European and other language families. The data suggest that the 2015 study by Haak et al. is flagrantly inconsistent with the archaeological record, the climatological record, the contemporary genetic date, and the ancient DNA data. R-M207 variation among Indo-European-speaking population was not shaped by a massive expansion of steppe nomads during the Bronze Age. Rather, the contemporary cross-linguistic distribution of R-M207 mutations in Eurasia and Africa was shaped by demographic processes that began at the end of the Last Glacial Maximum: expansions from Ice Age refugia; the demise of mega-fauna food resources; and population pressure associated with the Neolithic transition.

Section 12. Conclusions for R-M207.

For linguists, haplogroup R-M207 and its downstream variants provide especially useful makers for deciphering the prehistory of the Indo-European, Dravidian, Afro-Asiatic, and Nilo-Saharan language families as well as the Basque and Laal language isolates. One interesting observation is that language contact appears to partially explain the prehistoric expansion of Indo-European and Afro-Asiatic language families. R-M207 mutations were already in South Asia and Europe when Indo-European-speaking farmers arrived in these regions. Similarly, R-M207 was in northern Africa when Afro-Asiatic-speaking farmers arrived.

The *early farming dispersal hypothesis* remains a very robust model of Indo-European prehistory. This review of the evolutionary history of R-M207 and its downstream variants (Sections 1 to 11) demonstrates that palaeogenomic modeling has failed to undermine the hypothesis. Rather, palaeogenomic modeling produces conclusions that are flagrantly inconsistent with the archaeological record and other data sources. It is tempting to suggest that this problem could be cured by improving the statistical methods. However, autosomal markers may not be useful for deciphering human population history because of recombination. This reshuffling of the genetic "cards" seriously distorts evolutionary relationships and statistical methods cannot overcome this deficiency. Thus the message to linguists seems clear. Triangulated Y-chromosome based modeling is an alternative that is empirical, transparent and far more reliable.

Chapter 18: Recommendations, Observations and Future Research.

Section 1. Recommendations.

Chapters 2 to 17 explore the prehistory of language from a Y-chromosome perspective. The data suggest that triangulated Y-chromosome based models of language prehistory are highly reliable. The initial step in building such models is to identify informative Y-Chromosome mutations among contemporary populations for which language has a strong ethnic component. “Informative” generally means that a mutation has a moderate to high frequency (> 10%) among speakers of a specific language family. The next step in the model building process is to explain why a mutation attains a significant frequency. To resolve this question, data is extrapolated from phylogenetic relationships, ancient DNA, language relationships, the archaeological record, the paleo-climatological record, and other relevant sources.

Since Y-chromsome data are useful for modeling the prehistory of language, it would be in the best interest of linguists to encourage efforts that gather from populations that represent the full spectrum of linguistic diversity. Currently the amount of contemporary Y-chromosome data varies greatly from one region to the next. For example, European populations have been studied extensively. On the other hand, comparatively little data exists for Sub-Saharan Africa, Southeast Asia (especially Myanmar, Vietnam and Malaysia), the highlands region of New Guinea, and the indigenous peoples of North America (especially Alaska). Additionally, most of the data for aboriginal Australians has been taken from government databases that do not track group affiliation.

In some cases the paucity of data for a region may reflect the availability of funding that is available for population studies. European countries, for example, have the financial resources for these studies, whereas a country like Papua New Guinea may lack the resources. Moreover, it should also be emphasized that Native North American and aboriginal Australian populations have generally refused to participate in genetic studies because of an historical distrust of Europeans. These groups certainly represent a key component in understanding the evolution of language. Hopefully we can build alliances with them in the future.

Section 2. Miscellaneous Observations.

2.1. Overview.

The non-recombining region of the human acts as a “trap” which has successfully captured important demographic milestones that mark the evolutionary history of *Homo sapiens*. These data have greatly enhanced our understanding of the prehistory of language. Additionally, the data provide additional insight that explains, at least partially, the contemporary pattern of global linguistic diversity.

2.2. Migration, Climate, Language, and Cognition.

Prehistoric population expansions help to explain the contemporary pattern of language variation. Major expansions include the out-of-Africa exodus during Marine

Isotope Stage 5, the colonization of East Asia and Europe during Marine Isotope Stage 3, expansions from Ice Age refugia into the Americas during the late Pleistocene, and Neolithic agricultural expansions that occurred independently in several regions of the world. Interestingly, prehistoric human expansions were motivated, in part, by climate change. For example, less precipitation drove the out-of-Africa exodus. Warmer climatic conditions drove the human colonization of Europe, East Asia, and Australia during Marine Isotope Stage 3. Late Pleistocene deglaciation drove the human settlement of the Americas.

The survival of the animal and plant life of our world tends to be linked to a very limited ecological niche that offers opportunities and imposes constraints. As such, the survival of flora and fauna is sensitive to climate change. For example, an alligator in the Florida everglades is ill-equipped to survive on the frozen Arctic icepack. Similarly, a polar bear from the Arctic Circle is ill-equipped to survive in the everglades. This discussion of the general trend in ecology leads to an interesting observation. *Homo sapiens* are very adaptable. Climate driven prehistoric migrations pushed us into new biomes. Our cognitive abilities forged cultural adaptations that exploited the new opportunities that they offered and overcame the limitations that they imposed. For example, the mammoth hunters of Paleolithic Northern Eurasia perfected the hunting of large herbivores. As such, they successfully exploited the food resources of the tundra steppes. Furthermore, they utilized mammoth bones as fuels, which overcame a limitation of this region, cold climatic conditions.

The relationship between human cognition and language is a question that has arisen through the efforts to harness Y-chromsome data as a tool for linguistic research. It should be noted that the questions in not new, but rather the Sapir–Whorf hypothesis represents early efforts to explore this topic. This question surfaces once again because the data strongly suggest the following proposition: when Homo sapiens migrated out of Africa 100 thousand years ago, language was part of the social behavior of our species. Language, of course, presents especially strong evidence of the transition to modern human behavior. Taking this a step further, it seems that language is very much part of the evolutionary success of *Homo sapiens*. Language potentially facilitated collaborative problem solving that enabled us to survive in a wide variety of biomes.

2.3. Reproductive Success and Language.

Agriculture has drastically improved the reproductive success of *Homo sapiens*. From an evolutionary perspective, we found a survival strategy that supports far more people per square kilometer as compared to foraging. Rice, for example, supports over a billion people in East Asia. For linguists agriculture is a salient point because the Neolithic revolution drove rapid population growth. As a result language and farmers co-expanded in several regions of the world. These language-farmer expansions, in turn, partially explain the evolutionary history of the following language families: Indo-European, Niger-Congo, Afro-Asiatic, Uralic, Sino-Tibetan, Austro-Asiatic, Dravidian, Austronesian, Trans-New Guinea, and Arawak (Maipurean).

2.4. Human Evolutionary Adaptations.

Human evolutionary adaptations help to explain linguistic diversity. Admixture between Neanderthals and *Homo sapiens* may have strengthened the human immune system. The success of Tibeto-Burman languages stems from an evolutionary adaptation that enables

Tibetans to utilize the depleted oxygen level found on the Tibetan Plateau. They can overcome hypoxia and altitude sickness, significant health risks among those that inhabit this region. Similarly, Austronesians were able to farm the coastal areas of New Guinea because of an evolutionary adaptation that made them resistant to tropical splenomegaly syndrome, a massive and fatal enlargement of the spleen that occurs as the result of chronic exposure to malaria.

2.5. Language Contact.

It seems as though language contact theory has been underestimated as an explanation behind the global pattern of language variation. The evolutionary history of Germanic languages involves language contact between the Mesolithic populations of Scandinavia and the Neolithic populations of Central Europe. The evolution of Indo-Aryan also involves language contact between Neolithic farmers and Mesolithic hunter-gatherers. The Pygmies of the central African rainforest adopted the Niger-Congo languages of Bantu farmers. Papuans adopted the languages of Austronesian farmers and then expanded eastwards across the Pacific. In North and East Africa, during prehistoric times, Nilo-Saharan speaking populations shifted to Afro-Asiatic, and Afro-Asiatic populations shifted to Nilo-Saharan. The story of Austronesian languages in western Indonesia and Malaysia entails language shift from Austro-Asiatic to Austronesian. Finally, language contact best explains similarities found among the so-called Transeurasian languages.

Section 3. Unresolved Research Questions for the Future.

3.1. Introduction.

Early career linguists may want to specialize in one of the traditional sub-disciplines of linguistics, such as historical linguistics, or semantics, or morphology. Furthermore, the specialization should concentrate on gathering data from the historical record and field studies. Yet for those willing to take a risk, triangulated Y-chromosome modeling of language prehistory offers an exciting opportunity to set sail into an unexplored linguistic frontier. This section presents unresolved questions that represent innovative research opportunities for linguists of the future.

3.2. The Comparative Method and Language Prehistory.

One question involves the comparative method and the potential contribution provided by this linguistic tool for elucidating the prehistory of language. The linguist Lyle Campbell, for example, tends to employ a conservative application of the comparative method in an effort to classify language diversity. The linguist Robert Beekes, on the other hand, employs the comparative method to reconstruct the culture of "Indo-Europeans," a people that has never been documented either by the archaeological or historical record. A good working hypothesis might suggest that the comparative method is robust tool for language classification and not for cultural reconstruction.

3.3. Non-Linguistic Data and Language Classification.

Should we utilize non-linguistic data to classify languages? Controversial classifications such as Nilo-Saharan and Niger-Congo seem more robust when considers the genetic and anthropological data. Nilo-Saharan correlates well with the desertification of the Sahara and the rise of cattle herding subsistence in East Africa. Niger-Congo correlates well with the expansion of land agriculture from West Central Africa. On the other hand, classifications such as Sino-Tibetan seem less plausible when considers non-linguistic evidence. The so-called Chinese languages are linked to the success of rice agriculture in East Asia. Tibeto-Burman, on the other hand, evolved from the success of barley agriculture on the Tibetan and an evolutionary adaption among the Tibetan farmers that enables them to survive hypoxia.

3.4 Classifying Prehistoric Language Evolution.

How we can classify prehistoric language evolution? The early farming dispersal hypothesis is clearly a well-resolved model. However, the co-expansion of language and farming, as posited by the hypothesis, seems to be only one of several potential language evolution models. For example, Korean and Japanese represent the *in situ* co-evolution of agriculture and language. The distribution of Eyak-Athabaskan definitely follows a co-expansion of hunter-gatherers and language. Finally, linguistic diversity in Pacific Northwest (e.g. Tsimshian, Wakashan, and Salish) seems to have evolved along an *in situ* hunter-gatherer trajectory. As such, models of language prehistory await further refinement and classification.

3.5. Neolithic Revolution and the Leveling of Linguistic Diversity.

The special relationship between language variation and agriculture raises an interesting research question. Is there an inverse relationship between agriculture and language? In other words, did the Neolithic revolution level linguistic diversity? This question arises from the observation that linguistic diversity in South America has been difficult to classify. Compared to Eurasia and Africa, many of the South American languages are listed by *Ethnologue* as isolates or as unclassified. Additionally, linguistic diversity in South America consists of numerous small language families, whereas linguistic diversity in the Old World consists of comparatively fewer language families which in many cases, consist of hundreds of languages. As such, one could argue that linguistic distance is greater for New World languages than for Old World languages. Extending this argument further, this dichotomy may reflect that agriculture was practiced less intensively in prehistoric South America. On the other hand, the classification of indigenous language diversity in South America may reflect the availability of resources for historical linguistics. Alternatively, European colonization may have erased large sections of the linguistic map and as such, this has obscured linguistic relationships that facilitate language classification.

Section 4. Final Thoughts.

Interest in the prehistory of language has circulated within linguistic debate since the founding of our discipline over two hundred years ago. Nevertheless, among contemporary linguists some believe that the question of language prehistory is far too speculative, that it defies empirical analysis. This approach to linguistic research is problematic because the past explains the present. The more we know about the prehistory of language, the more we know about contemporary languages. Towards this goal some linguists have attempted to model the prehistory of language by using linguistic tools. These attempts, however, have rendered models of language that are sometimes clearly implausible, such as the correlation between the Basque isolate and the languages of the Caucasus region. Linguistic tools are also limited in achieving time depth. Triangulated Y-chromosome based modeling of language prehistory yields desperately needed models of language prehistory that are highly reliable. Moreover, we can drill much deeper into the prehistory of language.

Index of Languages

Bibliography

Aaris-Sorensen, Kim et al. 2007. “The Scandinavian reindeer (*Rangifer tarandus L.*) after the last glacial maximum: time, seasonality and human exploitation.” *Journal of Archaeological Science* 34(6): 914-923.

Abilev, Serikbai et al. 2012. “The Y-Chromosome C3 star-cluster attributed to Genghis Khan’s descendants is present at high frequency in the Kerey clan from Kazakhstan.” *Human Biology* 84(1): 79-89.

Abi-Rached, Laurent et al. 2011. “The shaping of modern human immune systems by multiregional admixture with archaic humans.” *Science* 334 (6052): 89-94.

Abramova et al. 2001. “The age of Upper Paleolithic sites in the middle Dnieper River basin of Eastern Europe.” *Radiocarbon* 43(2b): 1077-1084.

Abu-Amero, Khaled K. et al. 2009 . “Saudi Arabian Y-chromosome diversity and its relationship with nearby regions.” *BioMed Central Genetics* 10:59.

Ahn, Sung-Mo 2010. “The emergence of rice agriculture in Korea: archaeobotanical perspectives.” *Archaeological and Anthropological* Sciences 2: 89-98.

Aikens, C. Melvin and Takeru Akazawa 1996. “The Pleistocene-Holocene transition in Japan and adjacent northeast Asia.” In: *Humans at the end of the Ice Age*. Straus, Lawrence Guy et al. (Eds.). Plenum Press, New York.

Allen, Bryan 1992. “The geography of Papua New Guinea.” In: *Human Biology in Papua New Guinea: the Small Cosmos*. Edited by Robert D. Attenborough and Michael P. Alpers. Oxford; New York; Tokyo; Melbourne: Clarendon Press, pp. 36-66.

Allen, Jim and James F. O'Connell 2008. “Getting from Sunda to Sahul.” In*: Islands of Inquiry: Colonisation, Seafaring and the Archaeology of Maritime Landscapes*. Edited by Geoffrey Clark, Foss Leach, and Sue O’Connor. Canberra: ANU Press, pp. 31-46.

Allentoft, M. E. et al. 2015. “Population genomics of Bronze Age Eurasia.” *Nature* 522: 167-172.

Ammerman, A.J. and L.L Cavalli-Sforza 1984*. The Neolithic Transition and the Genetics of Populations in Europe*. Princeton: Princeton University Press.

Anthony, David W. 2007. *The Horse, the Wheel, and Language, How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World.* Princeton, N.J.; Woodstock: Princeton University Press.

Anthony, David 2008. “A new approach to language and archaeology: the Usatovo culture and the separation of Pre-Germanic. *Journal of Indo-European Studies* 36 (1/2): 1-51.

Anthony, David W. and Don Ringe 2015. “The Indo-European homeland from linguistic and archaeological perspectives.” Annual Review of Linguistics 1:199-219.

Arredi, Barbara et al. 2004. “A predominately Neolithic origin for Y-chromosomal DNA variation in North Africa.” *American Journal of Human Genetics* 75: 338-345.

Arrow, K.J., Panosian C., and Gelband H. (editors) 2004. *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance.* Washington (DC): Institute of Medicine (US) Committee on the Economics of Antimalarial Drugs. National Academies Press, pp. 125-135. https://www.ncbi.nlm.nih.gov/books/NBK215638/

Arunkumar, Ganesh Prasad et al. 2012. "Population differentiation of southern Indian male lineages correlates with agricultural expansions predating the caste system." *Public Library of Science One* 7(11): e50269.

Arunkumar, Ganesh Prasad et al. 2015. "A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west: late Neolithic expansion of O2a1-M95." *Journal of Systematics and Evolution* 53(6): 546-560.

Askarov, A. et al. 1992. "Pastoral and nomadic tribes at the beginning of the first millennium B.C." In: *History of Civilizations of Central Asia*. Vol. 1, the Dawn of Civilization: Earliest Times to 700B.C. Edited by A. H. Dani and V. M. Masson Paris: UNESCO Publishing, pp. 450-463.

Austerlitz, Robert 2009. "Uralic Languages." In: *The World's Major Languages*. Second Edition. Edited by Bernard Comrie. Oxon, UK; New York: Routledge, pp. 477-482.

Arvidsson, Stefan 2006. *Aryan Idols: Indo-European Mythology as Ideology and Science.* Chicago: University of Chicago Press.

Bahuchet, Serge 2012. "Changing language, remaining Pygmy." *Human Biology* 84 (1):11-43.

Bailey, Geoff et al 2007. "Coastal prehistory in the Southern Red Sea Basin, underwater archaeology and the Farasan Island." *Proceedings of the Seminar for Arabian Studies* 37: 1-16

Bailey, Robert C. et al. 1989. "Hunting and gathering in tropical rain forest: is it possible?" *American Anthropologist* 91: 59-82.

Balanovsky, O. et al. 2008. "Two sources of the Russian patrilineal heritage in their Eurasian context." American Journal of Human Genetics 82: 236-250.

Balanovsky et al. 2011. "Parallel evolution of genes and languages in the Caucasus region." *Molecular Biology and Evolution* 28(10): 2905-2920.

Balanovsky, Oleg et 2015. "Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y chromosome and reveals migrations of Iranic speakers." *Public Library of Science One* 10(4): e0122968.

Balanovsky, O. 2017. "Toward a consensus on SNP and STR mutation rates on the human Y-chromosome." *Human Genetics* 136:575-590.

Balanovsky, O. et al. 2017. "Genetic differentiation between upland and lowland populations shapes the Y chromosomal landscape of West Asia." *Human Genetics* 136: 437-450.

Balaresque, Patricia et al. 2010. "A predominantly Neolithic origins for European paternal lineages." *Public Library of Science Biology* 8(1): 1-9.

Barbieri, Chiara et al. 2012. "Contrasting maternal and paternal histories in the linguistic context of Burkina Faso." *Molecular Biology and Evolution* 29(4): 1213-1223.

Barbieri, Chiara et al. 2016. “Refining the Y chromosome phylogeny with southern African sequences.” *Human Genetics* 135: 541-553.

Barker, Graeme et al. 2007. “The ‘human revolution’ in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo).” *Journal of Human Evolution* 52: 243-261.

Bar-Yosef, Ofer 1998. “The Natufian culture in the Levant, threshold to the origins of agriculture.” *Evolutionary Anthropology* 6(5): 159-177.

Baskin, Leonid M. 1986. “Differences in the ecology and behavior of reindeer populations in the USSR.” *Rangifer* Special Issue No. 1: 333-340.

Baskin, Leonid M. 2003. “River crossings as principal points of human/reindeer relationship in Eurasia.” *Rangifer* Special Issue No. 14: 37-40.

Bates, J., C.A. Petriea, and R.N. Singh 2017. “Approaching rice domestication in South Asia: new evidence from Indus settlements in northern India.” *Journal of Archaeological Science* 78: 193-201.

Batini, Chiara et al. 2011. “Signatures of the preagricultural peopling processed in Sub-Saharan Africa as revealed by the phylogeography of early Y-chromosome lineages.” *Molecular Biology and Evolution* 28(9): 2603-2613.

Battaglia, Vincenza et al. 2009. “Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe.” *European Journal of Human Genetics* 17: 820-830.

Battaglia, Vincenza et al. 2013. “The first peopling of South America: new evidence from Y-chromosome haplogroup Q.” *Public Library of Science One* 8(8): e71390.

Bellwood, Peter 2005. *First farmers: the origins of agricultural societies*. Malden, MA; Oxford, UK; Victoria, Australia: Blackwell Publishing.

Benedict, Paul K. 1987. “Early MY/TB loan relationships.” *Linguistics of the Tibeto-Burman Area* 10(2): 12-21

Berger, Burkhard et al. 2013. “High resolution mapping of Y haplogroup G in Tyrol (Austria).” *Forensic Science International: Genetics* 7: 529-536.

Bergstrom, Anders et al. 2016. “Deep roots for Aboriginal Australian Y chromosomes.” *Current Biology* 26: 809-813.

Berniell-Lee, Gemma et al 2009. “Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages.” *Molecular Biology and Evolution* 26(7): 1581-1589.

Binney, Heather et al. 2016. “Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns.” *Quaternary Science Reviews* 157: 80-97.

Blench, Roger 2006. *The Niger-Saharan Macrophylum*. Cambridge, UK. Publisher: Author. http://www.rogerblench.info/Language/Nilo-Saharan/NS%20page.htm

Blench, Roger 2010. “Evidence for the Austronesian voyages in the Indian Ocean.” In: *The Global Origins and Development of Seafaring*. Edited by Atholl Anderson et al. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 239-248.

Blench, Roger 2013. “The Prehistory of the Daic- or Kra-Dai-speaking peoples and the hypothesis of an Austronesian Connection.” In: *Unearthing Southeast Asia’s Past: Selected Papers from the 12th International Conference of the European Association of Southeast Asian Archaeologists.* Edited by Marijke J. Klokke and Véronique Degroot. Singapore: NUS Press, pp. 3-15. http://www.jstor.org/stable/j.ctv1qv3nd.6.

Blench, Roger 2014**.** Linguistic and archaeological evidence for Berber prehistory. http://www.rogerblench.info/Language/Afroasiatic/AASOP.htm

Blockley, S.P.E. and R. Pinhasi 2011. “A revised chronology for the adoption of agriculture in the Southern Levant and the role of Late Glacial climatic change.” *Quaternary Science Reviews* 30: 98-108.

Blome, Margaret Whiting et al. “The environmental context for the origins of modern human diversity: A synthesis of regional variability in African climate 150,000-30,000 years ago.” *Journal of Human Evolution* 62(5): 563-592.

Blust, Robert 2013. *The Austronesian Languages*. Revised Edition. Canberra, Australia: Asia-Pacific Linguistics Research School of Pacific and Asian Studies. The Australian National University.

Bolnick, Deborah A. et al. 2004. “Problematic use of Greenberg’s linguistic classification of the Americas in studies of Native American genetic variation.*” American Journal of Human Genetics* 75: 519-523.

Bolnick, Deborah A. et al. 2006. “Asymmetric male and female genetic histories among Native Americans from Eastern North America.” *Molecular Biology and Evolution* 23(11): 2161-2174.

Bosch, Elena et al. 2001. “High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula.” *American Journal of Human Genetics* 68: 1019-1029.

Bostoen, Koen et al. 2015. “Middle to Late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of western central Africa.” *Current Anthropology* 56(3): 354-384.

Bourgeon, Lauriane, Ariane Burke and Thomas Higham 2017. “Earliest human presence in North America dated to the Last Glacial Maximum: new radiocarbon dates from Bluefish Caves,Canada.” *Public Library of Science One* 12(1): e0169486.

Bowler, James M. et al. 2003. “New ages for human occupation and climatic change at Lake Mungo, Australia.” *Nature* 421: 837-840.

Brackney, Noel C. 2007. *The Origins of Slavonic: Language Contact and Language Change*. Munich: LINCOM Europa.

Brown, Wesley M. 1980. “Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis.” *Proceedings of the National Academy of Sciences*. 77(6): 3605-3609.

Brunelli, Andrea et al. 2017. “Y chromosomal evidence on the origin of northern Thai people.” *Public Library of Science One* 12(7): e0181935.

Bruno, Maria C. 2006. “A morphological approach to documenting the domestication of Chenopodium in the Andes.” In: *Documenting Domestication: New Genetic and Archaeological Paradigms*. Edited by Melinda A. Zeder et al. Berkeley and Los Angeles: University of California Press, pp. 32-45.

Bučková, Jana et al. 2013. “Multiple and differentiated contributions to the male gene pool of pastoral and farmer populations of the African Sahel.” *American Journal of Physical Anthropology* 151: 10-21.

Buvit, Ian and Karisa Terry 2016. “Outside Beringia: why the Northeast Asian Upper Paleolithic record does not support a long standstill model.” *Paleo America* 2(4): 281-285.

Cadenas, Alicia M. et al. 2008. “Y-chromosome diversity characterizes the Gulf of Oman.” *European Journal of Human Genetics* 16: 374 - 386.

Cai, Xiaoyun et al. 2011. “Human migration through bottlenecks from Southeast Asia into East Asia during the last glacial maximum revealed by Y-chromosomes.” *Public Library of Science One* 6(8): e24282.

Campbell, Lyle 1997. *American Indian Languages: the Historical Linguistics of Native America*. New York: Oxford University Press.

Campbell, Lyle 2011. “The Dené –Yeniseian Connection (Kari and Potter, eds.).” Book review. *International Journal of American Linguistics* 77(3): 445-451.

Campbell, Lyle and Terrence Kaufman 1976. “A linguistic look at the Olmecs.” *American Antiquity* 41(1): 80-89.

Cann, Rebecca L., Mark Stoneking and Allan C. Wilson 1987. “Mitochondrial DNA and human evolution.” *Nature* 325: 31-36.

Capelli, C. et al. 2006. “Population structure in the Mediterranean Basin: a Y chromosome perspective.” Annals of Human Genetics 70: 207-225.

Capelli, Cristian et al. 2007. “Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter.” *Molecular Phylogenetics and Evolution* 44: 228-239.

Capredon, Melanie et al. 2013. “Tracing Arab-Islamic inheritance in Madagascar: study of the Y-chromosome and Mitochondrial DNA in the Antemoro.” *Public Library of Science On*e 8(11): e80932.

Capriles, Jose M. et al. 2019. “Persistent Early to Middle Holocene tropical foraging in southwestern Amazonia.” *Science Advances* 5: eaav5449.

Casanova, Myriam et al. 1985. “A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance.” *Science* 230(4732): 1403-1406.

Cassidy, Lara M. et al. 2016. “Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome.” *Proceedings of the National Academy of Sciences of the United States of America* 113(2): 368-373.

Cavalli-Sforza, L.L., Paolo Menozzi, and Alberto Piazza 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Cavalli-Sforza, Luigi Luca 2000. *Genes, Peoples and Languages*. Berkeley, CA; Los Angeles, CA: University of California Press.

Chaubey, Gyaneshwer et al. 2011. "Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture." *Molecular Biology and Evolution* 28(2): 1013-1024.

Chiaroni, Jacques et al. 2008. "Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent." *Antiquity* 82: 281-289.

Chiaroni, Jacques et al. 2009. "Y-chromosome diversity, human expansion, drift, and cultural evolution." *Proceeding of the National Academy of Sciences of the United States of America* 106(48): 20174-20179.

Chiaroni, Jacques et al. 2010. "The emergence of Y-chromosome haplogroup J1e among the Arabic-speaking populations." *European Journal of Human Genetics* 18: 348-353.

CIA World Factbook. https://www.cia.gov/library/publications/the-world-factbook/

Cinnioğlu, Cengiz et al. 2004. "Excavating Y-chromosome haplotype strata in Anatolia." *Human Genetics* 114: 127-14

Clark, Jeffrey T. and Kevin M. Kelly 1993. "Human genetics, paleoenvironments, and malaria: relationships and implications for the settlement of Oceania." *American Anthropologist* 95(3): 612-630.

Clark, Peter U. et al. 2009. "The Last Glacial Maximum." *Science* 325: 710-714.

Comrie, Bernard 2008. "Linguistic diversity in the Caucasus." *Annual Review of Anthropology* 37: 131-143

Cordaux, Richard et al. 2004. "Independent origins of Indian caste and tribal paternal lineages." *Current Biology* 14: 231-235.

Cox, Francis 2010. "History of the discovery of the malaria parasites and their vectors." *Parasites and Vectors* 3:5.

Cox, Murray P. et al. 2007. "A Polynesian motif on the Y chromosome: population structure in remote Oceania." *Human Biology* 79(5): 525-535.

Crowther, Alison et al. 2017. "Subsistence mosaics, forager-farmer interactions, and the transition to food production in eastern Africa." *Quaternary International* 489: 101-120.

Cruciani, Fulvio et al. 2004. "Phylogenetic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa." *American Journal of Human Genetics* 74: 1014-1022.

Cruciani, Fulvio et al. 2007. "Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosome haplogroups E-M78 and J-M12." *Molecular Biology and Evolution* 24(6): 1300-1311.

Cruciani, Fulvio et al. 2010. "Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages." *European Journal of Human Genetics* 18: 800-807.

Currat, Mathias and Laurent Excoffier 2011. "Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression." *Proceedings of the National Academy of Sciences of the United States of America* 108(37): 15129-15134.

D'Atanasio et al. 2018. "The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages." *Genome Biology* 19: 20.

Darwin, Charles, 1809-1882. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859.

Debnath, Monojit et al. 2011. "Y-chromosome haplogroup diversity in the sub-Himalayan Terai and Duars populations of East India." *Journal of Human Genetics* 56: 765-771.

Delfin, Frederick et al. 2011. "The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetics affinities of Negrito and non-Negrito groups." *European Journal of Human Genetics* 19: 224-230.

Delfin, Frederick et al. 2012. "Bridging Near and Remote Oceania: mtDNA and NRY variation in the Solomon Island." Molecular Biology and Evolution 29(2): 545-564.

Delfin, Frederick C. 2015. "The population history of the Philippines: a genetic overview." Philippine Studies: *Historical and Ethnographic Viewpoints* 63(4): 449-476.

Demeter, Fabrice et al. 2012. "Anatomically modern human in Southeast Asia (Laos) by 46 ka." *Proceedings of the National Academy of Sciences of the United States of America* 109 (36) 14375-14380

Denham, T.P. et al. 2003. "Origins of agriculture at Kuk Swamp in the highlands of New Guinea." *Science* 301: 189-193.

Di Cristofaro, Julie et al. 2013. "Afghan Hindu Kush: where Eurasian Sub-Continent gene flows converge." *Public Library of Science One* 8(10): e76748.

Diamond, Jared and Peter Bellwood 2003. "Farmers and their languages: the first expansions." *Science* 300: 597-603.

Diamond, Jared M. 2000. "Taiwan's gift to the world." *Nature* 403: 709-710.

Dillehay, Tom D. et al. 2015. "New archaeological evidence for an early human presence at Monte Verde, Chile." *Public Library of Science One* 10(11): e0141923.

Dixon, R.M.W. and Alexandra Y. Aikenvald 1999. "Introduction." In: *The Amazonian Languages*. Edited by R.M.W Dixon and Alexandra Y. Aikenvald. Cambridge, UK: Cambridge University Press, pp. 1-21.

Doi, Tran Tri 2012. "On the relationship between the Austroasiatic and Austronesian languages in Southeast Asia." *Vietnam National University Journal of Social Sciences* 28(5E): 35-39.

Dolukhanov, Pavel M. 2003. "Hunter-gathers of the Last Ice Age in Northern Eurasia." *Before Farming* 2003(2): 1-25

Dolukhanov, P. 2009. "The Mesolithic of the East European Plain." In: *The East European Plain on the Eve of Agriculture*. Edited by Pavel M. Dolukhanov, Graeme R. Sarson and Anvar M. Shukurov. BAR International Series 1964. Oxford, UK: Archaeopress, pp. 23-34.

Donohue, Mark and Tim Denham 2010. "Farming and language in Island Southeast Asia: reframing Austronesian history." *Current Anthropology* 51(2): 223-256.

Drake, Nick A. et al. 2011. "Ancient watercourses and biogeography of the Sahara explain the peopling of the desert." *Proceedings of the National Academy of Sciences of the United States of America* 108(2): 458-462.

Driem, George van 2005. "Tibeto-Burman vs Indo-Chinese: implications for population geneticists, archaeologists and prehistorians." In: *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*." Edited by Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas. London and New York: Routledge, pp. 81-106.

Driem, George van 2011. "Rice and the Austroasiatic and Hmong-Mien homelands." In: *Dynamics of Human Diversity*. Edited by N. J. Enfield. Canberra: Pacific Linguistics, Australian National University, pp. 361-390.

Driem, George van 2014. "Trans-Himalayan." In: *Trans-Himalayan Linguistics*. Edited by Nathan Hill and Thomas Owen-Smith. Berlin: Mouton de Gruyter, pp. 11-40.

Duggan, Ana et al. 2013. "Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers." *Public Library of Science One* 8(12): e83570.

Dumond, Don E. 2005. "The Arctic Small Tool Tradition in Southern Alaska." *Alaska Journal of Anthropology* 3(2): 67-78.

Dupuy, Berit Myhre et al. 2006. "Geographical heterogeneity of Y chromosomal lineage in Norway." *Forensic Science International* 164: 10-19.

Dyke, Arthur S. 2004. "An outline of North American deglaciation with emphasis on central and northern Canada." *Developments in Quaternary Sciences* 2(B): 373-424.

Ehret, Christopher et al. 2004. "The Origins of Afroasiatic." *Science* 306: 1680-1681.

Eldredge, Sandy and Bob Biek. "Ice Ages – What are they and what causes them?" *Utah Geological Survey*. https://geology.utah.gov/map-pub/survey-notes/glad-you-asked/ice-ages-what-are-they-and-what-causes-them/ Accessed September 28, 2017.

El-Sibai, Mirvat et al. 2009. "Geographical structure of the Y-Chromosomal genetic landscape of the Levant: a coastal-inland contrast." *Annals of Human Genetics* 73: 568-581.

Ennafaa, Hajer et al. 2011. "Mitochondrial and Y-chromosome microstructure in Tunisia." *Journal of Human Genetics* 56: 734-741.

Erlandson, John M., Madonna L. Moss and Matthew Des Lauriers 2008. "Life on the edge: early maritime cultures of the Pacific Coast of North America." *Quaternary Science Reviews* 27: 2232-2245.

Ethnologue 2014. 17th edition. Edited by Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2015. 18th edition. Edited by Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2016. 19th edition. Edited by M. Paul Lewis, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2017. 20th edition. Edited by Simons, Gary F. and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2018. 21st edition. Edited by Simons, Gary F. and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2019. 22nd edition. Edited by David M. Eberhard, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Fadhlaoui-Zid, Karima et al. 2011. "Genetic structure of Tunisian ethnic groups revealed by parental lineages." American Journal of Physical Anthropology 148: 271-280.

Fadhlaoui-Zid, Karima et al. 2013. "Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa." *Public Library of Science One* 8(11): e80293.

Fedorova, Sardana A. et al. 2013. "Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia." *BioMed Central Evolutionary Biology* 13:127.

Fehér, T. et al. 2015. "Y‑SNP L1034: limited genetic link between Mansi and Hungarian‑speaking populations." *Molecular Genetics and Genomics* 290: 377-386.

Fiedel, Stuart 2008. "Sudden deaths: the chronology of terminal Pleistocene megafaunal extinction." In: *American Megafaunal Extinctions at the End of the Pleistocene*. Edited by G. Haynes. Springer, pp. 21-37.

Filippo, Cesare de et al. 2011. "Y-chromosomal variation in Sub-Saharan Africa: Insights into the history of Niger-Congo groups." *Molecular Biology and Evolution* 28(3): 1255-1269.

Firasat, Sadaf et al. 2007. "Y-chromosomal evidence for a limited Greek contribution to the Pathan population of Pakistan." *European Journal of Human Genetics* 15: 121-126.

Fischer, Anders 2002. "Food for feasting? An evaluation of explanations for the neolithisation of Denmark and southern Sweden." In: The Neolithisation of Denmark: 150 Years of Debate. Eds. Anders Fischer and Kristian Kristiansen. Sheffield, England: J.R. Collins Publications, pp. 343-393.

Flegontov, Pavel et. al. 2016a. "Na-Dené populations descend from the Paleo-Eskimo migration into America." doi: https://doi.org/10.1101/074476

Flegontov, Pavel et al. 2016b. "Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry." *Scientific Reports* 6:20768.

Flegontov, Pavel et al. 2017. "Paleo-Eskimo genetic legacy across North America." doi: https://doi.org/10.1101/203018

Fortescue, Michael 2004. "How far west into Asia have Eskimo languages been spoken, and which ones?" *Etudes Inuit Studies* 28(2): 159-183.

Fortescue, Michael 2013. "North America: Eskimo-Aleut linguistic history." In: *The Global Prehistory of Human Migration.* Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 340-345.

Fortes-Lima, Cesar et al. 2015. "Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups." *Forensic Science International: Genetics* 19: 232–237.

Frachetti, Michael D. 2012. "Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia." *Current Anthropology* 53(1): 2-38.

Francalacci, Paolo et al. 2015. "Detection of Phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample." *BioMed Central Research Notes* 8:174

Fregel, Rosa et al. 2017. "Neolithization of North Africa involved the migration of people from both the Levant and Europe." bioRxiv preprint first posted online Sep. 21, 2017; doi: http://dx.doi.org/10.1101/191569

Friesen, T. Max 2013. "North America: Paleoeskimo and Inuit archaeology." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 346-353.

Frison, George C. 1998. "Paleoindian large mammal hunters on the plains of North America." *Proceedings of the National Academy of Sciences of the United States of America* 95: 14576-14583.

Fromm, Hans 1997. "Germanen in bronzezeitlichen Mittelschweden?" Finnisch-Ugrische Forschungen 54: 127-150.

Frumkin, Amos et al. 2011. "Possible paleohydrologic and paleoclimatic effects on hominin migration and occupation of the Levantine Middle Paleolithic." *Journal of Human Evolution* 60: 437-451.

Fu, Qiaomei et al. 2014. "Genome sequence of a 45,000-year-old modern human from western Siberia." *Nature* 514: 445-450.

Fu, Qiaomei et al. 2016. "The genetic history of Ice Age Europe." *Nature* 534: 200-205.

Fuller, Dorian Q. 2006. "Agricultural origins and frontiers in South Asia: a working synthesis." *Journal of World Prehistory* 20: 1-86.

Fuller, Dorian Q 2012. "Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures." *Rice* 4: 78-92.

Gaikwad, Sonali and VK Kashyap 2005. "Molecular insight into the genesis of ranked caste populations of western India based upon polymorphisms across non-recombinant and recombinant regions in genome." *Genome Biology* 6: P10.

Gan, Rui-Jing et al. 2008. "Pinghua population as an exception of Han Chinese's coherent genetic structure." *Journal of Human Genetics* 53: 303-313.

Gavashelishvili, Alexander and David Tarkhnishvili 2016. "Biomes and human distribution during the last Ice Age." *Global Ecology and Biogeography* 25: 563-574.

Gayà-Vidal, Magdalena et al. 2011. "mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: different stories and special genetic traits of the Andean Altiplano populations." *American Journal of Physical Anthropology* 145: 215-230.

Gayden, Tenzin et al. 2007. "The Himalayas as a directional barrier to gene flow." *American Journal of Human Genetics* 80: 884-894.

Gazi, Nurun Nahar et al. 2013. "Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal." *Public Library of Science One* 8(10): e75064.

Geppert, Maria et al. 2011. “Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America.” *Forensic Science International: Genetics* 5: 100-104.

Gillispie, Thomas E. 2018. An overview of Alaska's prehistoric cultures. Government publication. Office of History and Archeology Report 173. Anchorage, Alaska: Alaska Department of Natural Resources.

Gimbutas, Marija 1997. “The Fall and Transformation of Old Europe: Recapitulation 1993.” In: *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Eds. Marija Gimbutas; Miriam Robbins Dexter and Karlene Jones-Bley. Washington D.C.: Institute for Study of Man. 351-372.

Gomes, Verónica et al. 2010. “Digging deeper into East African human Y-chromosome lineages. *Human Genetics* 127: 603-613.

Gordon, Bryan 2003. “Rangifer and man: An ancient relationship.” *Rangifer*, Special Issue 14: 15-28.

Green, Richard E. et al. 2010. “A draft sequence of the Neanderthal genome.” *Science* 328: 710-722.

Greenberg, Joseph H. 1987. *Languages in the Americas*. Stanford, CA: Stanford University Press.

Grollemund, Rebecca et al. 2015. “Bantu expansion shows that habitat alters the route and pace of human dispersals.” *Proceedings of the National Academy of Sciences of the United States of America* 112(43): 13296–13301.

Groucutt, Huw S. et al. 2015. “Rethinking the dispersal of Homo sapiens out of Africa.” *Evolutionary Anthropology* 24:149-164.

Grugni, Viola et al. 2012. “Ancient migratory events in the Middle East: new clues from the Y-Chromosome variation of modern Iranians.” *Public Library of Science One* 7(7): e41252.

Grugni, Viola et al. 2019. “Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas.” *BioMed Central Biology* 17:3.

Güldemann, Tom and Mark Stoneking 2008. “A historical appraisal of clicks: a linguistic and genetic perspective.” *Annual Review of Anthropology* 37: 93-109.

Günther, Torsten et al. 2015. “Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques.” *Proceedings of the National Academy of Sciences of the United States of America* 112(38): 11917-11922.

Gusmão, Al. et al. 2008. “A perspective on the history of the Iberian gypsies provided by phylogenetic analysis of Y-chromosome lineages.” *Annals of Human Genetics* 72: 215-227.

Haak, Wolfgang et al. 2010. “Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities.” *Public Library of Science Biology* 8 (11).

Haak, Wolfgang et al. 2015. “Massive migration from the steppe was a source for Indo-European languages in Europe.” *Nature* 522: 207-211.

Haas R. et al. 2017. “Humans permanently occupied the Andean highlands by at least 7 ka.” *Royal Society Open Science* 4: 170331.

Haber, Marc et al. 2016. “Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations.” *American Journal of Human Genetics* 99: 1316–1324.

Hamilton, Marcus J. and Briggs Buchanan 2010. “Archaeological support for the three-stage expansion of modern humans across northeastern Eurasia and into the Americas.” *Public Library of Science One* 5(8): e12472.

Hammer, Michael F. 1994. “A recent insertion of an Alu element on the Y Chromosome is a useful marker for human population studies.” *Molecular Biology and Evolution* 11(5): 749-761.

Hammer, Michael F. et al. 2006. “Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y-chromosomes.” *Journal of Human Genetics* 51: 47-58.

Hardigan, Michael A. et al. 2017. “Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (46): E9999-E10008.

Hassan, Hisham Y. et al. 2008. “Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography and history.” *American Journal of Physical Anthropology* 137: 316-323.

He, J. and Guo, F. 2013. “Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning province, Northeast China.” *Forensic Science International-Genetics* 7: E84–E85.

He, Jun-Dong et al. 2012a. “Patrilineal perspective on the Austronesian diffusion in mainland Southeast Asia.” *Public Library of Science One* 7(5): e36437.

Heckenberger, Michael 2013. “Amazonia: archaeology.” In: *The Global Prehistory of Human Migration.* Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 392-400.

Heggarty, Paul and David Beresford-Jones 2010. “Agriculture and language dispersals: limitations, refinements, and an Andean exception?” *Current Anthropology* 51(2): 163-193.

Heggarty, Paul and David Beresford-Jones 2013. “Andes: linguistic history.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 401-409.

Heizer, Robert F. 1944. *Aconite Poison Whaling in Asia and America: an Aleutian Transfer to the New World.* U.S. government document. Anthropological Papers, No. 24. Smithsonian Institution, Bureau of Ethnology, Bulletin 133.

Henn, Brenna M. et al. 2008. “Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa.” *Proceedings of the National Academy of Sciences of the United States of America* 105(31): 10693-10698.

Henn, Brenna M. et al. 2011. “Hunter-gatherer genomic diversity suggests a southern African origin for modern humans.” *Proceedings of the National Academy of Sciences of the United States of America* 108(13): 5154–5162.

Herrera, Kristian J. et al. 2012. “Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists.” *European Journal of Human Genetics* 20: 313-320.

Hershkovitz, Israel et al. 2015. “Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans.” *Nature* 520: 216-219.

Hetzron, Robert 2009. Afro-Asiatic languages. In: *The World’s Major Languages. Second Edition*. Bernard Comrie, Ed. Oxon, UK; New York: Routledge. 545-550.

Higham, Charles (2002). “Languages and farming dispersals: Austroasiatic languages and rice cultivation.” In: *Examining the Farming/Language Dispersal Hypothesis*. Edited by Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, pp. 223-232.

Hill, A. V. S. et al. 1985. “Melanesians and Polynesians share a unique α-thalassemia mutation.” *American Journal of Human Genetics* 37: 571-580.

Hill, Jane H. 2001. “Proto-Uto-Aztecan: a community of cultivators in Central Mexico?” *American Anthropologist* 103(4): 913-934.

Hirschenfeld, Ludwik and Hanka Hirschenfeld 1919. “Serological differences between the blood of different races.” *The Lancet* 194 (5016): 675-679.

Hoffecker, John F. 2009. “The spread of modern humans in Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (38):16040–16045.

Hoffecker, John F., Scott A. Elias BS and Dennis H. O’Rourke 2014. “Out of Beringia?” *Science* 343: 979-980.

Holmes, Charles E. 2011. “The Beringian and transitional periods in Alaska: technology of the East Beringian tradition as viewed from Swan Point.” In: *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia*. Edited by Ted E. Goebel and Ian Buvit. Center for the Study of the First Americans. College Station, Texas: A&M University Press, pp. 179-191.

Holst Pellekaan, Sheila van 2013. “Genetic evidence for the colonization of Australia.” *Quaternary International* 285: 44-56.

Horborg, Alf 2005. “Ethnogenesis, regional integration, and ecology in prehistoric Amazonia.” *Current Anthology* 16(4): 589-620.

Horsburgh, K. Ann and Mark D. McCoy 2017. “Dispersal, isolation, and interaction in the islands of Polynesia: a critical review of archaeological and genetic evidence.” *Diversity* 9:37.

Housley, R.A. et al. 1997. “Radiocarbon evidence for the late glacial human recolonisation of Northern Europe.” *Proceedings of the Prehistoric Society* 63: 25-54.

Hovhannisyan, Anahit et al. 2014. “Different waves and directions of Neolithic migrations in the Armenian Highland.” *Investigative Genetics* 5:15.

Hu, Kang et al. 2015. “The dichotomy structure of Y chromosome Haplogroup N.” https://arxiv.org/abs/1504.06463

Huang,Yun Zhi et al. 2017. “Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407.” *Molecular Genetics and Genomics* 293(3): 657-663.

Huang, Yun‑Zhi et al. 2018. “Dispersals of the Siberian Y‑chromosome haplogroup Q in Eurasia.” *Molecular Genetics and Genomics* 293:107-117.

Hublin, Jean-Jacques et al. 2017. “New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens.” *Nature* 546: 289-292.

Hudjashov, Georgi et al. 2007. “Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis.” *Proceeding of the National Academy of Sciences of the United States of America* 104 (21): 8726-8730.

Hudson, Mark J. 2013. “Japan: archaeology.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 224-229.

Hudson, Mark J. 2017. “The Ryukyu islands and the northern frontier of prehistoric Austronesian settlement.” In: *New Perspectives in Southeast Asian and Pacific Prehistory*. Edited by Philip J. Piper, Hirofumi Matsumura and David Bulbeck. Acton, Australia: Australian National University Press, pp. 189-200.

Hung, Hsiao-chun and Mike T. Carson 2014. “Foragers, fishers and farmers: origins of the Taiwanese Neolithic.” *Antiquity* 88: 1115-1131.

Ilumae, Anne-Mai et al. 2016. “Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families.” *American Journal of Human Genetics* 99: 163-173.

International Society of Genetic Genealogy. http://www.isogg.org.

Janhunen, Juha 2003a. “Para-Mongolic.” In: *The Mongolic Languages*. Juha Janhunen (Ed.). Routledge: London. 391-402.

Janhunen, Juha 2003b. “Written Mongol.” In: *The Mongolic Languages*. Juha Janhunen (Ed.). Routledge: London. 30-56.

Jessen, Catherine A. et al. 2015. “Early Maglemosian culture in the Preboreal landscape: Archaeology and vegetation from the earliest Mesolithic site in Denmark at Lundby Mose, Sjælland.” *Quaternary International* 378: 73-87.

Jobling, Mark A. and Chris Tyler-Smith 2003. “The human Y chromosome: an evolutionary marker comes of age.” *Nature Reviews: Genetics* 4: 598-612.

Jochim, Michael et al. 1999. “The Magdalenian colonization of southern Germany.” *American Anthropologist* 101:129-142.

Jones, Eppie R. et al. 2015. “Upper Palaeolithic genomes of modern Eurasians.” *Nature Communications* 6: 8912.

Jones, Eppie R. et al. 2017. “The Neolithic transition in the Baltic was not driven by admixture with early European farmers.” *Current Biology* 27: 576–582.

Jota, Marilza S. et al. 2016. “New native South American Y chromosome lineages.” *Journal of Human Genetics* 61(7): 593-603.

Kahlke, Ralf-Dietrich 2015. “The maximum geographic extension of Late Pleistocene Mammuthus primigenius (Proboscidea, Mammalia) and its limiting factors.” *Quaternary International* 379: 147-154.

Kane, Daniel 1989. *The Sino-Jurchen Vocabulary of the Bureau of Interpreters*. Bloomington, Indiana: Indiana University Institute for Inner Asian Studies.

Kang, Longli et al. 2012. “Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas.” *Annals of Human Genetics* 76: 92-99.

Karachanak, Sena et al. 2013. “Y-chromosome diversity in modern Bulgarians: new clues about their ancestry.” *Public Library of Science One* 8(3): e56779.

Karafet, Tatiana M. et al. 2002. “High levels of Y-chromosome differentiation among Native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life.” *Human Biology* 74(6): 761-789.

Karafet, Tatiana et al. 2008. “New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree.” *Genome Research* 18: 830-838.

Karafet, Tatiana M. et al. 2010. “Major east-west division underlies Y-Chromosome stratification across Indonesia.” *Molecular Biology and Evolution* 27(8): 1833-1844.

Karafet, Tatiana M. et al. 2015. “Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia.” *European Journal of Human Genetics* 23: 369-373.

Karafet, Tatiana et al. 2016. “Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan.” *Journal of Human Genetics* (2016) 61: 181-191.

Karlsson, Andreas O. et al. 2006. “Y-chromosome diversity in Sweden - a long-time perspective.” *European Journal of Human Genetics* 14: 963-970.

Karmin, Monika et al. 2015. “A recent bottleneck of Y chromosome diversity coincides with a global change in culture.” *Genome Research* 25: 459-466.

Kayser, Manfred et al. 2003. “Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea.” *American Journal of Human Genetics* 72: 281-302.

Kayser, Manfred et al. 2006. “Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific.” *Molecular Biology and Evolution* 23(11): 2234-2244.

Kayser, Manfred et al. 2008. “The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia.” *Molecular Biology and Evolution* 25(7): 1362-1374.

Keegan, William 2013. “Caribbean Islands: archaeology.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp 376-383.

Kemp, Brian M. et al. 2007. “Genetic analysis of Early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas.” *American Journal of Physical Anthropology* 132: 605-621.

Kemp, Brian M. et al. 2010. “Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in Southwest and Mesoamerica.” *Proceedings of the National Academy of Sciences of the United States of America* 107(15): 6759-6764.

Khurana, Priyanka et al. 2014. “Y Chromosome haplogroup distribution in Indo-European speaking tribes of Gujarat, Western India.” *Public Library of Science One* 9(3): e90414.

Kim, Nam-Kil 2009. “Korean.” In: *The World’s Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge, pp. 765-779.

Kim, Seung-Og 2015. “Recent developments and debates in Korean prehistoric archaeology.” *Asian Perspectives* 54(1): 11-30.

Kim, Soon-Hee et al. 2011. “High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea.” *Investigative Genetics* 2:10

King, R.J. et al. 2008. “Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic.” *Annals of Human Genetics* 72: 205-214.

Kistler, Logan et al. 2018. “Multiproxy evidence highlights a complex evolutionary legacy of maize in South America.” *Science* 362, 1309-1313.

Kivisild, T. et al. 2003. “The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.” *American Journal of Human Genetics* 72: 313-332.

Knight, Alec et al. 2003. “African Y chromosome and mtDNA divergence provides insight into the history of the click languages.” *Current Biology* 13: 464-473.

Kolpaschikov, Leonid et al. 2015. “The role of harvest, predators, and socio-political environment in the dynamics of the Taimyr wild reindeer herd with some lessons for North America.” *Ecology and Society* 20(1): 9.

Kornfilt, Jaklin 2009. “Turkish and the Turkic languages.” In: *The World’s Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge 519-544.

Kosaka, Ryuichi. 2002. “On the affiliation of Miao-Yao and Kadai: can we posit the Miao-Dai family.” *Mon-Khmer Studies* 32: 71-100.

Kumar, Vikrant et al. 2007. “Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations.” *BioMed Central Evolutionary Biology* 7: 47.

Kuper, Rudolph and Stefan Kroepelin 2006. “Climate-controlled Holocene occupation in the Sahara: motor of Africa’s evolution.” *Science* 313: 803-807.

Kusuma, Pradiptajati et al. 2015. “Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations.” *BioMed Central Genomics* 16: 191.

Kuzmin, Yaroslav V. 2008. “Siberia at the Last Glacial Maximum: environment and archaeology.” *Journal of Archaeological Research* 16: 163-221.

Kwon, So Yeun et al. 2015. "Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups." *Forensic Science International Genetics* 19: 42-46.

Lacan, Marie et al. 2011. "Ancient DNA suggests the leading role played by men in the Neolithic dissemination." *Proceeding of the National Academy of Sciences of the United States of America* 108(45): 18255-18259.

Lacau, Harlette et al. 2012. "Afghanistan from a Y-chromosome perspective." *European Journal of Human Genetics* 20: 1063-1070.

Lahn, Bruce T. et al. 2001. "The human Y chromosome in light of evolution." *Nature Reviews Genetics* 2: 207-216.

Laitinen, Virpi et al. 2002. "Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers." *Human Heredity* 53:68-78.

LaPolla, Randy J. 2001. "The role of migration and language contact in the development of the Sino-Tibetan Language Family." In: *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*. Edited by R. M. W. Dixon and A. Y. Aikhenvald. Oxford: Oxford University Press, pp. 225-254.

LaPolla, Randy J. 2013. "Eastern Asia: Sino-Tibetan linguistic history." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 204-208.

LaPolla, Randy J. 2016. "Problems with the arguments for recasting Sino-Tibetan as 'Trans-Himalayan.'" *Linguistics of the Tibeto-Burman Area* 39:(2): 282-297.

Lappalainen, Tuuli et al. 2006. "Regional differences among the Finns: a Y-chromosome perspective." *Gene* 376: 207-215.

Lappalainen, T. et al. 2008. "Migration waves to the Baltic Sea region." *Annals of Human Genetics* 72: 337-348.

Larmuseau, M.H.D. et al 2014. "Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations." *Forensic Science International: Genetics* 9: 179-185.

Lazaridis, Iosif et al. 2014. "Ancient human genomes suggest three ancestral populations for present-day Europeans." *Nature* 513: 409-413

Lazaridis, Iosif et al. 2016. "Genomic insights into the origin of farming in the ancient Near East." *Nature* 25: 536 (7617): 419-424.

LeBlanc, Steven A. 2013. "Mesoamerica and the southwestern United States: archaeology." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp 369-375.

Lee, Eun Young et al. 2014. "Analysis of 22 Y chromosomal STR haplotypes and Y haplogroup distribution in Pathans of Pakistan." *Forensic Science International: Genetics* 11: 111-116.

Lee, Ki-Moon and S. Robert Ramsey 2011. *A History of Korean Language*. Cambridge, UK: Cambridge University Press.

Lewis J.P. et al. 2016. “The shellfish enigma across the Mesolithic-Neolithic transition in southern Scandinavia.” *Quaternary Science Reviews* 151: 315-320.

Li, Hui et al. 2007. “Y chromosomes of prehistoric people along the Yangtze River.” *Human Genetics* 122: 383-388.

Li, Dongna et al. 2008a. “Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia.” *Public Library of Science One* 3(5): e2168.

Li, Dongna et al. 2010. “Genetic origin of Kadai-speaking Gelong people on Hainan island viewed from Y chromosomes.” *Journal of Human Genetics* 55: 462-468.

Li, Dong-Na et al. 2013. “Substitution of Hainan indigenous genetic lineage in the Utsat people, exiles of the Champa kingdom.” *Journal of Systematics and Evolution* 51(3): 287-294.

Liu, Wu et al. 2015. “The earliest unequivocally modern humans in southern China.” *Nature* 526: 696–699.

Loikala, Paula. 1977. “The oldest Germanic loanword in Finnish and their contribution to German philology.” *Studi italiani di linguistica teorica ed applicata*. 6 (1-2): 223-250.

López-Parra, A.M. et al. 2009. “In search of the pre- and post-Neolithic genetic substrates in Iberia: Evidence from Y-chromosome in Pyrenean populations.” *Annals of Human Genetics* 73: 42-53.

Loy, Dorothy E. et al. 2018. “Evolutionary history of human Plasmodium vivax revealed by genome-wide analyses of related ape parasites.” *Proceedings of the National Academy of Sciences of the United States of America* 115(36): E8450–E8459.

Luis, J.R. et al. 2004. “The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.” *American Journal of Human Genetics* 74: 532-544.

Mailhammer, Robert 2007. The Germanic Strong Verbs: *Foundations and Development of a New System*. Berlin; New York: Mouton de Gruyter.

Malhi, Ripan Singh et al. 2008. “Distribution of Y chromosomes among native north Americans: a study of Athabaskan population history.” *American Journal of Physical Anthropology* 137: 412-424.

Malyarchuk, Boris et al. 2011. “Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a.” *Journal of Human Genetics* 56: 583-588.

Mann, Daniel H. et al. 2015. “Life and extinction of megafauna in the ice-age Arctic.” *Proceedings of the National Academy of Sciences of the United States* 112(46): 14301-14306.

Marcheco-Teruel, Beatriz et al. 2014. “Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers.” *Public Library of Science Genetics* 10(7): e1004488.

Martins, Haidé et al. 2015. “Radiocarbon Dating the Beginning of the Neolithic in Iberia: New Results, New Problems.” *Journal of Mediterranean Archaeology* 28(1): 105-131.

Matson, R. G. and M. P. R. Magne 2013. “North America: Na Dené /Athapaskan archeology and linguistics.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 333-339.

McEvoy, Brian et al. 2004. “The longue durée of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic façade of Europe.” *American Journal of Human Genetics* 75: 693-702.

Mellars, Paul 2006. “Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia.” *Science* 313: 796-800.

Mendez, Fernando L. et al. 2011. “Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa.” *Human Biology* 83(1): 39-53.

Mendez, Fernando L. et al. 2013. “An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree.” *American Society of Human Genetics* 92: 454-459.

Mengoni Gonalons, Guillermo L. and Hugo D. Yaco Baccio 2006. “The domestication of South American camelids: a view from the South-Central Andes.” In: *Documenting Domestication: New Genetic and Archaeological Paradigms*. Edited by Melinda A.Zeder et al. Berkeley and Los Angeles: University of California Press, pp. 228-244.

Merrill, William L. 2009. “The diffusion of maize to the southwestern United States and its impact.” *Proceedings of the National Academy of Sciences of the United States* 106(50): 21019-21029.

Mezzavilla, Massimo et al. 2015. “Insights into the origin of rare haplogroup C3* Y chromosomes in South America from high-density autosomal SNP genotyping.” *Forensic Science International: Genetics* 15: 115-120.

Militarev, Alexander 2002. “The prehistory of a dispersal: the Proto-Afrasian (Afroasiatic) farming lexicon.” In: *Examining the Farming/Language Dispersal Hypothesis*. Edited by Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, pp. 135-150.

Mirabal, Sheyla et al. 2012. “Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga.” *Gene* 492: 339-348.

Mirov, N. T. 1945. “Notes on the domestication of reindeer.” *American Anthropologist* 47: 393-408.

Misra, Virendra Nath 2001. “Prehistoric human colonization of India.” *Journal of Biosciences* 26(4):491-531.

Mitchell, Peter 2010. “Genetics and southern African prehistory: an archaeological view.” *Journal of Anthropological Sciences* 88: 73-92.

Mona, Stefano et al. 2007. “Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis.” *Molecular Biology and Evolution* 24(11): 2546-2555.

Mona, Stefano et al. 2009. “Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis.” *Molecular Biology and Evolution* 26(8): 1865-1877.

Mona, Stefano et al. 2011. Corrigendum. *Molecular Biology and Evolution* 26(8):1865-1877 28(8): 2419-2420.

Mondal, Mayukh et al. 2017. “Y‑chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese.” *Human Genetics* 136(5):499-510.

Montano, Valeria et al. 2011. “The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa.” *Molecular Ecology* 20: 2693-2708.

Morein, Eugene 2008. “Evidence for declines in human population densities during the early Upper Paleolithic in western Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 105(1): 48-53.

Morelli, Laura et al. 2010. “A comparison of Y-chromosome variation in Sardinia and Anatolia is more consistent with cultural rather than demic diffusion of agriculture.” *Public Library of Science One* 5(4): e10419.

Moreno-Mayar, J. Víctor et al. 2018. “Early human dispersals within the Americas.” *Science* 362 (6419): eaav2621.

Moss, Madonna L. and Jon M. Erlandson 1995. “Reflections on North American Pacific coast prehistory.” *Journal of World Prehistory* 9(1): 1-45.

Müller, Ulrich C. et al. 2011. “The role of climate in the spread of modern humans into Europe.” *Quaternary Science Reviews* 30: 273-279.

Mulligan, Connie J. and Emoke J.E. Szathmary 2017. “The peopling of Americas and the origin of the Beringian occupation model.” *American Journal of Physical Anthropology* 162: 403-408.

Myres, Natalie M. et al. 2011. “A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe.” *European Journal of Human Genetics* 19: 95-101

Nagle, Nano et al. 2016a. “Antiquity and diversity of aboriginal Australian Y-chromosomes.” *American Journal of Physical Anthropology* 159: 367-381.

Nagle, Nano et al. 2016b. “Mitochondrial DNA diversity of present-day aboriginal Australians and implications for human evolution in Oceania.” *Journal of Human Genetics* 62: 343-353.

Naitoh, Sae et al. 2013. “Assignment of Y-chromosomal SNPs found in Japanese population to Y-chromosomal haplogroup tree.” *Journal of Human Genetics* 58: 195-201.

Narasimhan, Vagheesh M. et al. 2019. “The formation of human populations in South and Central Asia.” *Science* 365: eaat7487.

Neves, Walter Alves, Diogo Meyer, and Hector Mario Pucciarella 1996. “Early skeletal remains and the peopling of the Americas.” *Revista de Anthropologia* 39(2): 121-139.

Ning, Chao et al. 2016. “Refined phylogenetic structure of an abundant East Asian Y-chromosomal haplogroup O*-M134.” *European Journal of Human Genetics* 24: 307-309.

Noble, William and Iain Davidson 1991. “The emergence of modern human behavior, language and its archaeology.” *Man* 26 (2): 223-253.

Norquest, Peter K. 2007. *A Phonological Reconstruction of Proto-Hlai*. Dissertation. University of Arizona, 2007.

Olalde, Inigo et al 2018. “The beaker phenomenon and the genomic transformation of Northwest Europe.” *Nature* 555: 190-196.

Olofsson, Jill Katharina et al. 2015. “Peopling of the North Circumpolar Region – insights from Y Chromosome STR and SNP typing of Greenlanders.” *Public Library of Science One* 10(1): e0116573.

Olsen, Kenneth M. and Barbara A. Schaal 1999. “Evidence on the origins of cassava: Manihot esculenta.” *Proceedings of the National Academy of Sciences of the United States of America* 96: 5586-5591.

Ono, Akira et al. 2002. “Radiocarbon dates and archaeology of the Late Pleistocene in the Japanese Islands.” *Radiocarbon* 44(2): 477-494.

Oppenheimer, Stephen 2012. “Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 770-784.

Ostapirat, Weera 2018. “Macrophyletic trees of East Asian languages re-examined.” *Senri Ethnological Studies* 98: 107-121.

Ottoni, Claudio et al. 2011. “Deep into the roots of the Libyan Tuareg: a genetic survey of their paternal heritage.” *American Journal of Physical Anthropology* 145: 118-124.

Oven, Mannis van and Manfred Kayser 2008. “Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation.” *Human Mutation* 30(2): E386-E394.

Oven, Mannis van et al. 2014. “Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome.” *Human Mutation* 35:187-191.

Pakendorf, Brigitte et al. 2006. “Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of the Yakuts.” *Human Genetics* 120: 334-353.

Pakendorf, Brigitte et al. 2007. “Mating patterns amongst Siberian Reindeer herders: inferences from mtDNA and Y-chromosomal analyses.” *American Journal of Physical Anthropology* 133: 1013-1027.

Pamjav, Horolma et al. 2011. “Genetic structure of the parental lineage of the Roma people.” *American Journal of Physical Anthropology* 145: 21-29.

Pamjav, Horolma et al. 2012. “New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1.” *American Journal of Physical Anthropology* 149: 611-615.

Park, Myung Jin et al. 2012. “Understanding the Y chromosome variation in Korea—relevance of combined haplogroup and haplotype analyses.” *International Journal of Legal Medicine* 126: 589–599.

Parton, Ash et al. 2015. “Orbital-scale climate variability in Arabia as a potential motor for human dispersals.” *Quaternary International* 382 82-97.

Pavlov, B.M. et al. 1994. “Population dynamics of the Taimyr reindeer population.” *Rangifer* Special Issue No. 9.

Pawley, Andrew 2005. “The chequered career of the Trans New Guinea hypothesis: recent research and its implications.” In: *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*. Edited by Andrew Pawley et al. Canberra, Australia: Pacific Linguistics, pp. 67-107.

Pedersen, Holger 1967. *The Discovery of Language: Linguistic Science in the Nineteenth Century.* Translated by John Webster Spargo. Bloomington: Indiana University Press.

Peng, Min-Sheng et al. 2014. “Retrieving Y-chromosomal haplogroup tree GWAS data.” *European Journal of Human Genetics* 22: 1046-1050.

Pereira, Luísa et al. 2010. “Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel.” *European Journal of Human Genetics* 18: 915-923.

Petrejčíková, Eva et al. 2009. “Y-haplogroup frequencies in the Slovak Romany population.” *Anthropological Science* 117(2): 89-94.

Pfeifer, Sebastian J. et al. 2019. “Mammoth ivory was the most suitable osseous raw material for the production of Late Pleistocene big game projectile points.” *Scientific Reports* 9:2303.

Pimenoff, Ville N. 2008. “Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers.” *European Journal of Human Genetics* 16: 1254-1264.

Pinotti, Thomaz et al. 2019. “Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American founders.” Current Biology 29: 149-157.

Pitulko, V. V. and P. A. Nikolskiy 2012. “The extinction of the woolly mammoth and the archaeological record in Northeastern Asia.” *World Archaeology* 44(1): 21-42.

Pitulko, V.V. et al. 2004. “The Yana RHS site: humans in the Arctic before the Last Glacial Maximum.” *Science* 303: 52-56.

Pitulko, Vladimir V. et al. 2016. “Early presence in the Arctic: evidence from 45,000-year-old mammoth remains.” *Science* 351(6270): 260-263.

Platt, Daniel E. et al. 2017. “Mapping post-glacial expansions: the peopling of Southwest Asia.” *Scientific Reports* 7: 40338.

Poetsch, Micaela et al. 2013. “Determination of population origin: a comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example.” *European Journal of Human Genetics* 21: 1423-1428.

Pope, Kevin O. and John E. Terrell 2008. “Environmental setting of human migrations in the circum-Pacific region.” *Journal of Biogeography* 35: 1-21.

Potter, Ben A. et al. 2014. “New insights into Eastern Beringian mortuary behavior: a terminal Pleistocene double infant burial at Upward Sun River.” *Proceedings of the National Academy of Sciences of the United States of America* 111(48): 17060-17065.

Potter, Ben A. et al. 2018. “Current evidence allows multiple models for the peopling of the Americas.” *Science Advances* 4: eaat5473.

Poznik, G. David et al. 2016. “Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences.” *Nature Genetics* 48(6): 593-600.

Pringle, Heather 2006. *The Master Plan: Himmler's Scholars and the Holocaust.* New York: Hyperion.

Puzachenko, A.Yu. and A.K. Markova 2019. “Evolution of mammal species composition and species richness during the Late Pleistocene - Holocene transition in Europe: a general view at the regional scale.” *Quaternary International* in press.

Qamar, Raheel et al. 2002. “Y-chromosomal DNA variation in Pakistan.” *American Journal of Human Genetics* 70: 1107-1124.

Qi, Xuebin et al. 2013. “Genetic evidence of Paleolithic colonization and Neolithic expansion of modern humans on the Tibetan Plateau.” *Molecular Biology and Evolution* 30(8): 1761-1778.

Quintana-Murci, Luís et al. 2008. “Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.” *Proceedings of the National Academy of Sciences of the United States of America* 105(5): 1596-1601.

Raghavan, Maanasa et al. 2014. “Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans.” *Nature* 505: 87-91.

Rai, Niraj et al. 2012. “The phylogeography of Y-chromosome haplogroup H1a1a-M82 reveals the likely Indian origin of the European Romani populations.*”* *Public Library of Science One* 7(11): e 48477.

Rasmussen, Morten et al. 2010. “Ancient human genome sequence of an extinct Palaeo-Eskimo.” *Nature* 463: 757-762.

Rasmussen, Morten et al. 2014. “The genome of a Late Pleistocene human from a Clovis burial site in western Montana.” *Nature* 506: 225-229.

Rasmussen, Morten et al. 2015. “The ancestry and affiliations of Kennewick Man.” *Nature* 523: 455-458.

Reardon, Sara 2017. “Navajo Nation reconsidered ban on genetic research.” *Nature* 550: 165-166.

Redd, Alan J. et al. 2002. “Gene flow from the Indian Subcontinent to Australia: evidence from the Y Chromosome.” *Current Biology* 12: 673-677.

Regueiro, Maria et al. 2012. “*High levels of Paleolithic Y-chromosome lineages characterize Serbia*.” *Gene* 498: 59-67.

Regueiro, Maria et al. 2013. “On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists.” *American Journal of Physical Anthropology* 150: 333-348.

Regueiro, Maria et al 2015. “From Arabia to Iberia: a Y chromosome perspective.” *Gene* 564: 141-152.

Reich, David et al. 2011. “Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania.” *American Journal of Human Genetics* 89: 516–528.

Renfrew, Colin 1987. *Archaeology and Language: the Puzzle of Indo-European Origins*. Cambridge; New York: Cambridge University Press.

Ricankova, Vera Pavelkova et al. 2014. “Ecological structure of recent and Last Glacial mammalian faunas in Northern Eurasia: the case of Altai-Sayan refugium.” *Public Library of Science One* 9(1): e85056.

Robbeets, Martine 2008. The historical comparison of Japanese, Korean and the Trans-Eurasian languages. *Rivista degli studi orientali* 81:(1-4): 261-287.

Robbeets, Martine 2017a. “Proto-Trans-Eurasian: Where and When?” *Man in India* 97(1): 19-46.

Robbeets, Martine 2017b. “Austronesian influence and Transeurasian ancestry in Japanese.” *Language Dynamics and Change* 7: 210-251.

Rocha, Jorge and Anne-Maria Fehn 2016. “Genetics and demographic history of the Bantu.” In: Wiley eLS Online Library.
http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0022892/full

Røed, Knut H. et al. 2008. “Genetic analyses reveal independent domestication origins of Eurasian reindeer.” *Proceedings of the Royal Society B* 275: 1849-1855.

Roewer, Lutz et al. 2013. “Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in Native South Americans.” *Public Library of Science Genetics* 9(4): e1003460.

Roosevelt, A.C. et al. 1996. “Paleo-Indian cave dwellers in the Amazon: the peopling of the Americas.” *Science* 272: (5260): 373-384.

Rootsi, Siri et al. 2004. “Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe.” *American Journal of Human Genetics* 75: 128-137.

Rootsi, Siiri et al. 2007. “A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe.” *European Journal of Human Genetics* 15: 204-211.

Rootsi, Siiri et al. 2012. “Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus.” *European Journal of Human Genetics* 20: 1275-1282.

Rosa, Alexandra et al. 2007. “Y-chromosome diversity in the population of Guinea-Bissau: a multiethnic perspective.” *BioMed Central Evolutionary Biology* 7: 124.

Roullier, Caroline et al. 2013. “Disentangling the origins of cultivated sweet potato.” *Public Library of Science One* 8(5): e62707.

Rowold, Daine J. et al. 2016. “On the Bantu expansion.” *Gene* 593: 48-57.

Ruhlen, Merritt 1998. “The origin of Na-Dené.” *Proceedings of the National Academy of Sciences of the United States of America* 95: 13994-13996.

Ryan, William B.F. et al 2003. “A catastrophic flooding of the Black Sea.” *Annual Review of Earth and Planetary Sciences*. 31: 525-554.

Ryan, William B.F. et al. 1997. “An abrupt drowning of the Black Sea shelf.” *Marine Geology* 138: 119-126.

Sagart, Laurent 2004. “The higher phylogeny of Austronesian and the position of Tai-Kadai.” *Oceanic Linguistics* 43(2): 411-444.

Sahoo, Sanghamitra et al. 2006. “A prehistory of Indian Y-chromosomes: evaluating demic diffusion scenarios.” *Proceedings of the National Academy of Sciences of the United States of America* 103(3): 843-848.

Sanchez, J.J. et al. 2004. “Y chromosome SNP haplogroups in Danes, Greenlanders and Somalis.” *International Congress Series* 1261: 347-349.

Sarac, Jelena et al. 2016. “Genetic Heritage of Croatians in the Southeastern European Gene Pool - Y Chromosome Analysis of the Croatian Continental and Island Population.” *American Journal of Human Biology* 28:837–845.

Sato, Youichi et al. 2014. “Overview of genetic variation in the Y chromosome of modern Japanese males.” *Anthropological Science* 122(3): 131-136.

Savelle, James M. and Nobuhiro Kishigami 2013. “Anthropological research on whaling: prehistoric, historic and current contexts.” *Senri Ethnological Studies* 84: 1-48.

Schapper, Antoinette 2017. “Farming and the Trans-New Guinea family: a consideration.” In: *Language Dispersal beyond Farming*. Edited by Martine Robbeets and Alexander Savelyev. John Benjamins Publishing Company, pp. 155-181.

Scheib, C.L. et al. 2018. “Ancient human parallel lineages within North America contributed to a coastal expansion.” *Science* 360: 1024-1027.

Schirmer, Alfred and Walther Mitzka 1969. *Deutsche Wortkunde* 6th edition. Berlin: Walter de Gruyter.

Schleicher, August 1863. *Die Darwinsche Theories und die Sprachwissenschaft: Offenes Sendsschreiben an Herrn Dr. Ernst Häcket*. Weimar: Hermann Böhlau.

Schurr, Theodore G. et al. 2012. “Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska.” *American Journal of Physical Anthropology* 148: 422-435.

Scozzari, Rosaria et al. 2012. “Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree.” *Public Library of Science One* 7(11): e49170.

Semino, Ornella et al. 1996. “A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers.” *American Journal of Human Genetics* 59: 964-968.

Semino, Ornella et al 2004. “Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean area.” *American Journal of Human Genetics* 74: 1023-1034.

Sengupta, Sanghamitra et al. 2006. “Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveals minor genetic influence of central Asian pastoralists.” *American Journal of Human Genetics* 78: 201-221.

Serdyuk, Natalia 2005. “The history of mammalian communities and paleography of Altai mountains in the Paleolithic.” *Paleontological Journal* 39 (Suppl. 6): S645–S821.

Shi, Hong et al. 2008. “Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations.” *BMC Biology* 6: 45.

Shi, Hong et al. 2013. “Genetic evidence of an East Asian origin and Paleolithic northward migration of Y-chromosome haplogroup N.” *Public Library of Science One* 8(6): e66102.

Shibatani, Masayoshi 2009. “Japanese.” In: *The World’s Major Languages*. Second edition. Edited by Bernard Comrie. Oxon, UK; New York: Routledge, pp. 741-763.

Shimoji, Michinori 2010. “Ryukyuan languages: an introduction.” In: An Introduction to Ryukyuan Languages. Edited by Michinori Shimoji and Thomas Pellard. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, pp. 1-13.

Sidwell, Paul 2010. “The Austroasiatic central riverine hypothesis.” *Journal of Language Relationship* 4: 117-134.

Sidwell, Paul 2013. “Southeast Asian mainland: linguistic history.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 259-268.

Siiräinen, Ari 2003. “The Stone and Broze Ages.” In: *The Cambridge History of Scandinavia*, Vol. 1. Knut Helle (Ed). Cambridge University Press, pp. 43-59.

Sikora, Martin et al. 2017. “Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers.” *Science* 3:358: 659-662.

Sikora, Martin et al. 2018. “The population history of northeastern Siberia since the Pleistocene.” bioRxiv preprint. doi: https://doi.org/10.1101/448829

Singh, Sakshi et al. 2016. “Dissecting the influence of Neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup.” *Scientific Reports* 6: 19157.

Skoglund, Pontus et al. 2015. “Genetic evidence for two founding populations of the Americas.” *Nature* 525: 104-110.

Smalley, John and Michael Blake 2003. “Sweet beginnings: stalk sugar and the domestication of maize.” *Current Anthropology* 44(5): 675-703.

Snow, Dean 2013. “Eastern North America: archeology and linguistics.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 354-361.

Solé-Morata, Neus et al. 2017. “Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ.” *Scientific Reports* 7: 7341.

Sommer, Robert S. et al. 2014. “Range dynamics of the reindeer in Europe during the last 25,000 years.” *Journal of Biogeography* 41: 298-306.

Sondaar, Paul et al. 1995. “The human colonization of Sardinia: a late Pleistocene human fossil from Corbeddu cave.” *Comptes Rendus de l'Académie des Sciences Paris* 320: 145-150.

Spengler, Robert et al. 2014. “Early agriculture and crop transmission among Bronze Age mobile pastoralists of Central Eurasia.” *Proceedings of the Royal Society B* 281: 20133382.

Spooner, David M . et al. 2005. “A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (41): 14694–14699.

Sporrong, Ulf 2003. “The Scandinavian landscape and its resources.” In: *The Cambridge History of Scandinavia*, vol. 1. Edited by Knut Helle. *Cambridge University Press*, pp. 15-42.

St. Clair, Michael R. 2012. *Germanic Origins from the Perspective of the Y-chromosome.* Ph.D. dissertation. http://oskicat.berkeley.edu/record=b20247897~S1

St. Clair, Michael R. 2014. “Population genetics and the humanities: Crimean Goths and contemporary European Romani.” *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*. 19(1): 135-142.

Stevens, Chris J. and Dorian Q. Fuller 2017. “The spread of agriculture in eastern Asia: archaeological bases for hypothetical farmer/language dispersals*.” Language Dynamics and Change* 7: 152-186.

Stewart, J. R. and C. B. Stringer 2012. “Human evolution out of Africa: the role of refugia and climate change.” *Science* 1317-1321.

Stoneking, Mark and Frederick Delfin 2010. “The human genetic history of East Asia: weaving a complex tapestry.” *Current Biology* 20: R188-R193.

Stoneking, Mark et al. 1990. “Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124: 717-733.

Straus, Lawrence Guy, David J. Meltzer and Ted Goebel 2005. “Ice Age Atlantis? Exploring the Solutrean-Clovis ‘connection.’” *World Archaeology* 37(4): 507-532.

Summerhayes, Glenn R. and Atholl Anderson 2009. “An Austronesian presence in southern Japan: early occupation in the Yaeyama Islands.” *Bulletin of the Indo-Pacific Prehistory Association* 29: 76-91.

Szecsenyi-Nagy, Anna et al. 2015. “Tracing the genetic origin of Europe’s first farmers reveals insights into their social organization.” *Proceedings of the Royal Society B*. 282: 20150339.

Tamang, Rakesh et al. 2018. “Reconstructing the demographic history of the Himalayan and adjoining populations.” *Human Genetics* 137(2): 129-139.

Tambets, Kristina et al. 2004. “The western and eastern roots of the Saami - the story of genetic ‘outliers’ told by the mitochondrial DNA and Y Chromosome.” *American Journal of Human Genetics* 74: 661-682.

Tamm, Erika et al. 2007. “Beringian standstill and spread of Native American founders.” *Public Library of Science One* 2(9): e829.

Taylor, R.E., David Glenn Smith and John R Southon 2001. “The Kennewick skeleton: chronological and biomolecular contexts.” *Radiocarbon* 43(2B): 965-976.

Tcherenkov, Lev and Stéphane Laederich 2004. *The Rroma: otherwise known as Gypsies, Gitanos, Gyphtoi, Tsiganes, Ţigani, Çingene, Zigeuner, Bohémiens, Travellers, Fahrende, etc.* Schwabe Verlag.

Thangaraj, Kumarasamy et al. 2003. “Genetic affinities of the Andaman Islanders, a vanishing human population.” *Current Biology* 13: 86-93.

Thanseem, Ismail et al. 2006. “Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA.” *BioMed Central Genetics* 7: 42.

Thurgood, Graham 1994. “Tai-Kadai and Austronesian: the nature of the historical relationship.” *Oceanic Linguistics* 33(2): 345-368.

Tishkoff, Sarah A. et al. 2007. “History of click speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation.” *Molecular Biology and Evolution* 24(10): 2180-2195.

Tofanelli, Sergio et al. 2009a. “On the origins and admixture of Malagasy: new evidence from high resolution analyses of paternal and maternal lineages.” *Molecular Biology and Evolution* 26(9): 2109-2124.

Tofanelli, Sergio et al. 2009b. “J1-M267 Y lineage marks climate-driven pre-historical human displacements.” *European Journal of Human Genetics* 17: 1520-1524.

Tranter, Nicholas 2012. “Introduction: typology and area in Japan and Korea.” *The languages of Japan and Korea*. Edited by Nicolas Tranter. Oxon, UK: Routledge, pp. 3-23.

Trask, R.L. 1996. *Historical Linguistics*. London: Arnold.

Trejaut, Jean A. et al. 2014. “Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia.” *BioMed Central Genetics* 15: 77.

Tremayne, Andrew 2015. “New Evidence for the timing of Arctic Small Tool Tradition coastal settlement in northwest Alaska.” *Alaska Journal of Anthropology* 13(1): 1-18.

Triki-Fendri, Soumaya et al. 2015. “Paternal lineages in Libya inferred from Y-chromosome haplogroups.” *American Journal of Physical Anthropology* 57: 242–251.

Trinkaus, Erik et al. 2003. “An early modern human from the Peştera cu Oase, Romania.” *Proceedings of the National Academy of Sciences of the United States of America* 100(20): 11231–11236.

Trivedi, R. et al. 2008. “Genetic imprints of Pleistocene origins of Indian populations: a comprehensive phylogenetic sketch of Indian Y-chromosomes.” *International Journal of Human Genetics* 8(1-2) 97-118.

Trombetta, Beniamino et al. 2015. “Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent.” *Genome Biology and Evolution* 7(7): 1940–1950.

Tumonggor, Meryanne K. et al. 2014. “Isolation, contact and social behavior shaped genetic diversity in West Timor.” *Journal of Human Genetics* 59: 494-503.

Underhill, Peter A. and Toomas Kivisild 2007. “Use of Y-Chromosome and Mitochondrial DNA population structure in tracing human migrations.” *Annual Review of Genetics* 41: 539-564.

Underhill, Peter A. et al. 1996. “A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history.” *Proceedings of the National Academy of Sciences of the United States of America* 93: 196-200.

Underhill, Peter A. et al. 2000. “Y chromosome variation and the history of human populations.” *Nature Genetics* 26: 358-361.

Underhill, P.A. et al. 2001. “The phylogeography of Y chromosome binary haplogroups and the origins of modern human populations.” Annals of Human Genetics 65: 43-62.

Underhill, Peter A. et al. 2007. “New phylogenetic relationships for Y-chromosome haplogroup I.” In: *Rethinking the Human Revolution*. Edited by P. Mellars et al. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 33-42.

Underhill, Peter A. et al. 2015. “The phylogenetic and geographic structure of Y-chromosome haplogroup R1a.” *European Journal of Human Genetics* 23: 124-131.

Valverde, Laura et al. 2016. “European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia.” *European Journal of Human Genetics* 24: 437-441.

Velichko, A.A. et al. 2009. “Initial human settlement of East European Plain.” In: *The East European Plain on the Eve of Agriculture*. Edited by Pavel M. Dolukhanov, Graeme R. Sarson and Anvar M. Shukurov. BAR International Series 1964. Oxford, UK: Archaeopress, pp. 17-21.

Vennemann, Theo 2000. “Zur Entstehung des Germanischen.” *Sprachwissenschaft* 25(3): 233-269.

Villalba-Mouco, Vanessa et al, 2019. “Survival of Late Pleistocene hunter-gatherer ancestry in the Iberian Peninsula.” *Current Biology* 29(7): 1169-1177.

Völgyi, Antonia et al. 2008. “Haplogroup distribution of Hungarian population and the largest minority group.” Forensic Science International Genetics Supplement Series 1: 383-385.

Voskarides, Konstantinos et al. 2016. “Y-chromosome phylogeographic analysis of the Greek-Cypriot population reveals elements consistent with Neolithic and Bronze Age settlements.” *Investigative Genetics* 7:1.

Wang, Chuan Chao et al. 2013a. “Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611.” *Journal of Systematics and Evolution* 51(3): 280-286.

Wang, Chuan-Chao and Hui Li 2013b. “Inferring human history in East Asia from Y chromosomes.” *Investigative Genetics* 4:11.

Wang, Chuan-Chao et al. 2014. “Genetic structure of Qiangic populations residing in the Western Sichuan Corridor.” *Public Library of Science One* 9(8): e103772.

Waterman, John T. 1970. *Perspectives in Linguistics*. Chicago: University of Chicago Press.

Waters, Michael R. et al 2011. “Pre-Clovis mastodon hunting 13,800 years ago at the Manis Site Washington.” Science 334: 351-353.

Wei, Lan-Hai and Hui Li 2017. “Fuyan human of 120–80 kya cannot challenge the Out-of-Africa theory for modern human dispersal.” *Science Bulletin* 62: 316-318.

Wei, Lan-Hai et al. 2017a. “Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia.” *Public Library of Science One* 5:12(4): e0175080.

Wei, Lan-Hai et al. 2017b. “Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations.” *Journal of Human Genetics* (2017) 62, 915-918.

Wei, Lan-Hai et al. 2018. “Paternal origin of Paleo-Indians in Siberia: insights from Y-chromosome sequences.” *European Journal of Human Genetics* 26: 1687-1696.

Welsch, Robert L and Adam M Levine 2008. “New Guinea and Melanesia.” In: *Encyclopedia of Archaeology,* vol. 3. Edited by D. M. Pearsall. New York: Elsevier/Academic Press, pp. 1738-1747.

White, Tim D. et al. 2003. “Pleistocene Homo sapiens from Middle Awash, Ethiopia.” *Nature* 423: 742-747.

Whitney, Bronwen et al. 2014. “Pre-Columbian raised-field agriculture and land use in the Bolivian Amazon.” *The Holocene* 24(2): 231-241.

Williams, Jack F. 2003. “Who are the Taiwanese? Taiwan in the Chinese diaspora.” In: *The Chinese Diaspora*. *Space, Place, Mobility and Identity*. Edited by Laurence J.C. Ma and Carolyn Cartier. Rowman and Littlefield Publishers, pp. 163-189.

Wilson, Samuel M. 2007. The Archaeology of the Caribbean. Cambridge, UK: Cambridge University Press, pp. 59-136.

Winters, Clyde 2010. “Y-Chromosome evidence of an African origin of Dravidian agriculture.” *International Journal of Genetics and Molecular Biology* 2(3): 30-33.

Wood, Elizabeth T. 2005. “Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biases demographic processes.” *European Journal of Human Genetics* 13: 867-876.

Wu, Tianyi, and Bengt Kayser 2006. “High altitude adaptation in Tibetans.” *High Altitude Medicine and Biology* 7(3): 193-208.

Xue, Yali et al. 2006. “Male demography in East Asia: a north-south contrast in human population expansion times.” *Genetics* 172: 2413-2439.

Yan, Shi et al. 2014. “Y chromosomes of 40% of Chinese descend from three Neolithic super-grandfathers.” *Public Library of Science One* 9(8): e105691.

Yang, Jian et al. 2017. “Genetic signatures of high-altitude adaptations in Tibetans.” *Proceedings of the National Academy of Sciences of the United States of America* 114(16): 4189-4194.

YCC 2002. “A nomenclature system for the tree of human Y-chromosome binary haplogroups.” *Genome Research* 12: 339-348.

Yunusbayev, Bayazit et al. 2012. “The Caucasus as an asymmetric semipermeable barrier to ancient human migrations.” *Molecular Biology and Evolution* 29(1): 359-365.

Zalloua, Pierre A. et al. 2008. “Y-chromosome diversity in Lebanon is structured by recent historical events.” *The American Journal of Human Genetics* 82: 873-882.

Zegura, Stephen L. et al. 2004. “High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas.” *Molecular Biology and Evolution* 21(1): 164-175.

Zerjal, Tatiana et al. 1997. “Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis.” *American Journal of Human Genetics* 60: 1174-1183.

Zerjal, Tatiana et al. 2003. “The genetic legacy of the Mongols.” *American Journal of Human Genetics* 72: 717-721.

Zhabagin, Maxat et al. 2017. “The connection of the genetic, cultural and geographic landscapes of Transoxiana.” *Scientific Reports* 7: 3085.

Zhang Chi and Hsiao-chun Hung 2008. “The Neolithic of Southern China – origin, development, and dispersal.” *Asian Perspectives* 47(2): 299-329.

Zhang Chi and Hsiao-chun Hung 2010. “The emergence of agriculture in southern China.” *Antiquity* 84: 11-25.

Zhang, Chi and Hsiao-chun Hung 2012. “Later hunter-gatherers in southern China, 18000–3000 BC.” *Antiquity* 86: 11-29.

Zhang, Chi and Hung Hsiao-chun 2013. “Eastern Asia: archaeology.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 209-216.

Zhang, Dong Ju et al. 2016. “History and possible mechanisms of prehistoric human migration to the Tibetan Plateau.” *Science China Earth Science* 59(9): 1765-1778.

Zhang, Xiaoming et al. 2015. “Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent.” *Scientific Reports* 5:154 86.

Zhao, Qing et al. 2010. “Gene flow between Zhuang and Han populations in the China-Vietnam borderland.” *Journal of Human Genetics* 55: 774-776.

Zhao, Zhijun 2011. “The origins of agriculture: new data, new ideas.” *Current Anthropology* 52(S4): S295-S306.

Zhong, Hua et al. 2010. “Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia.” *Journal of Human Genetics* 55: 428-435.

Zhong, Hua et al. 2011. “Extended Y-chromosome investigations suggests postglacial migrations of modern humans into East Asia via the northern route.” *Molecular Biology and Evolution* 28(1): 717-727.